

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number  
**WO 02/42329 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/47

Boulevard, Apartment 260, San Jose, CA 95128 (US).  
LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San  
Jose, CA 95117 (US).

(21) International Application Number: PCT/US01/45177

(22) International Filing Date:  
20 November 2001 (20.11.2001)

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell  
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(26) Publication Language: English

(30) Priority Data:  
60/252,189 21 November 2000 (21.11.2000) US

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier application:

US 60/252,189 (CIP)  
Filed on 21 November 2000 (21.11.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): DI-  
ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard,  
South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): SALCEDA, Susana  
[AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136  
(US). MACINA, Roberto, A. [AR/US]; 4118 Crescendo  
Avenue, San Jose, CA 95136 (US). RECIPON, Herve  
[FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115  
(US). SUN, Yongming [US/US]; 869 S. Winchester

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 02/42329 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate tissue, identifying prostate tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

## COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application  
5 Serial No. 60/252,189 filed November 21, 2000, which is herein incorporated by  
reference in its entirety.

### FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and  
10 polypeptides present in normal and neoplastic prostate cells, including fragments,  
variants and derivatives of the nucleic acids and polypeptides. The present invention also  
relates to antibodies to the polypeptides of the invention, as well as agonists and  
antagonists of the polypeptides of the invention. The invention also relates to  
compositions comprising the nucleic acids, polypeptides, antibodies, variants,  
15 derivatives, agonists and antagonists of the invention and methods for the use of these  
compositions. These uses include identifying, diagnosing, monitoring, staging, imaging  
and treating prostate cancer and non-cancerous disease states in prostate tissue,  
identifying prostate tissue and monitoring and identifying and/or designing agonists and  
antagonists of polypeptides of the invention. The uses also include gene therapy,  
20 production of transgenic animals and cells, and production of engineered prostate tissue  
for treatment and research.

### BACKGROUND OF THE INVENTION

Prostate cancer is the most prevalent cancer in men and is the second leading  
cause of death from cancer among males in the United States. AJCC Cancer Staging  
25 Handbook 203 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998); Walter J. Burdette, Cancer:  
Etiology, Diagnosis, and Treatment 147 (1998). In 1999, it was estimated that 37,000  
men in the United States would die as result of prostate cancer. Elizabeth A. Platz et al.,  
& Edward Giovannucci, Epidemiology of and Risk Factors for Prostate Cancer, in  
Management of Prostate Cancer 21 (Eric A Klein, ed. 2000). Cancer of the prostate  
30 typically occurs in older males, with a median age of 74 years for clinical diagnosis.

Burdette, *supra* at 147. A man's risk of being diagnosed with invasive prostate cancer in his lifetime is one in six. Platz et al., *supra* at 21.

Although our understanding of the etiology of prostate cancer is incomplete, the results of extensive research in this area point to a combination of age, genetic and  
5 environmental/dietary factors. Platz et al., *supra* at 19; Burdette, *supra* at 147; Steven K. Clinton, *Diet and Nutrition in Prostate Cancer Prevention and Therapy*, in Prostate Cancer: A Multidisciplinary Guide 246-269 (Philip W. Kantoff et al. eds. 1997).

Broadly speaking, genetic risk factors predisposing one to prostate cancer include race and a family history of the disease. Platz et al., *supra* at 19, 28-29, 32-34. Aside from  
10 these generalities, a deeper understanding of the genetic basis of prostate cancer has remained elusive. Considerable research has been directed to studying the link between prostate cancer, androgens, and androgen regulation, as androgens play a crucial role in prostate growth and differentiation. Meena Augustus et al., *Molecular Genetics and Markers of Progression*, in Management of Prostate Cancer 59 (Eric A Klein ed. 2000).  
15 While a number of studies have concluded that prostate tumor development is linked to elevated levels of circulating androgen (e.g., testosterone and dihydrotestosterone), the genetic determinants of these levels remain unknown. Platz et al., *supra* at 29-30.

Several studies have explored a possible link between prostate cancer and the androgen receptor (AR) gene, the gene product of which mediates the molecular and  
20 cellular effects of testosterone and dihydrotestosterone in tissues responsive to androgens. *Id.* at 30. Differences in the number of certain trinucleotide repeats in exon 1, the region involved in transactivational control, have been of particular interest. Augustus et al., *supra* at 60. For example, these studies have revealed that as the number of CAG repeats decreases the transactivation ability of the gene product increases, as  
25 does the risk of prostate cancer. Platz et al., *supra* at 30-31. Other research has focused on the  $\alpha$ -reductase Type 2 gene, the gene which codes for the enzyme that converts testosterone into dihydrotestosterone. *Id.* at 30. Dihydrotestosterone has greater affinity for the AR than testosterone, resulting in increased transactivation of genes responsive to androgens. *Id.* While studies have reported differences among the races in the length of  
30 a TA dinucleotide repeat in the 3' untranslated region, no link has been established between the length of that repeat and prostate cancer. *Id.*

Interestingly, while *ras* gene mutations are implicated in numerous other cancers, such mutations appear not to play a significant role in prostate cancer, at least among Caucasian males. Augustus, *supra* at 52.

Environmental/dietary risk factors which may increase the risk of prostate cancer include intake of saturated fat and calcium. Platz et al., *supra* at 19, 25-26. Conversely, intake of selenium, vitamin E and tomato products (which contain the carotenoid lycopene) apparently decrease that risk. *Id.* at 19, 26-28 The impact of physical activity, cigarette smoking, and alcohol consumption on prostate cancer is unclear. Platz et al., *supra* at 23-25.

Periodic screening for prostate cancer is most effectively performed by digital rectal examination (DRE) of the prostate, in conjunction with determination of the serum level of prostate-specific antigen (PSA). Burdette, *supra* at 148. While the merits of such screening are the subject of considerable debate, Jerome P. Richie & Irving D. Kaplan, *Screening for Prostate Cancer: The Horns of a Dilemma*, in Prostate Cancer: A Multidisciplinary Guide 1-10 (Philip W. Kantoff et al. eds. 1997), the American Cancer Society and American Urological Association recommend that both of these tests be performed annually on men 50 years or older with a life expectancy of at least 10 years, and younger men at high risk for prostate cancer. Ian M. Thompson & John Foley, *Screening for Prostate Cancer*, in Management of Prostate Cancer 71 (Eric A Klein ed. 2000). If necessary, these screening methods may be followed by additional tests, including biopsy, ultrasonic imaging, computerized tomography, and magnetic resonance imaging. Christopher A. Haas & Martin I. Resnick, *Trends in Diagnosis, Biopsy, and Imaging*, in Management of Prostate Cancer 89-98 (Eric A Klein ed. 2000); Burdette, *supra* at 148.

Once the diagnosis of prostate cancer has been made, treatment decisions for the individual are typically linked to the stage of prostate cancer present in that individual, as well as his age and overall health. Burdette, *supra* at 151. One preferred classification system for staging prostate cancer was developed by the American Urological Association (AUA). *Id.* at 148. The AUA classification system divides prostate tumors into four broad stages, A to D, which are in turn accompanied by a number of smaller substages. Burdette, *supra* at 152-153; Anthony V. D'Amico et al., *The Staging of*



*Prostate Cancer*, in Prostate Cancer: A Multidisciplinary Guide 41 (Philip W. Kantoff et al. eds. 1997).

Stage A prostate cancer refers to the presence of microscopic cancer within the prostate gland. D'Amico, *supra* at 41. This stage is comprised of two substages: A1, which involves less than four well-differentiated cancer foci within the prostate, and A2, which involves greater than three well-differentiated cancer foci or alternatively, moderately to poorly differentiated foci within the prostate. Burdette, *supra* at 152; D'Amico, *supra* at 41. Treatment for stage A1 preferentially involves following PSA levels and periodic DRE. Burdette, *supra* at 151. Should PSA levels rise, preferred treatments include radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.*

Stage B prostate cancer is characterized by the presence of a palpable lump within the prostate. Burdette, *supra* at 152-53; D'Amico, *supra* at 41. This stage is comprised of three substages: B1, in which the lump is less than 2 cm and is contained in one lobe of the prostate; B2, in which the lump is greater than 2 cm yet is still contained within one lobe; and B3, in which the lump has spread to both lobes. Burdette, *supra*, at 152-53. For stages B1 and B2, the treatment again involves radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.* at 151. In stage B3, radical prostatectomy is employed if the cancer is well-differentiated and PSA levels are below 15 ng/mL; otherwise, external beam radiation is the chosen treatment option. *Id.*

Stage C prostate cancer involves a substantial cancer mass accompanied by extraprostatic extension. Burdette, *supra* at 153; D'Amico, *supra* at 41. Like stage A prostate cancer, Stage C is comprised of two substages: substage C1, in which the tumor is relatively minimal, with minor prostatic extension, and substage C2, in which the tumor is large and bulky, with major prostatic extension. *Id.* The treatment of choice for both substages is external beam radiation. Burdette, *supra* at 151.

The fourth and final stage of prostate cancer, Stage D, describes the extent to which the cancer has metastasized. Burdette, *supra* at 153; D'Amico, *supra* at 41. This stage is comprised of four substages: (1) D0, in which acid phosphatase levels are

persistently high, (2) D1, in which only the pelvic lymph nodes have been invaded, (3) D2, in which the lymph nodes above the aortic bifurcation have been invaded, with or without distant metastasis, and (4) D3, in which the metastasis progresses despite intense hormonal therapy. *Id.* Treatment at this stage may involve hormonal therapy, chemotherapy, and removal of one or both testes. Burdette, *supra* at 151.

Despite the need for accurate staging of prostate cancer, current staging methodology is limited. The wide variety of biological behavior displayed by neoplasms of the prostate has resulted in considerable difficulty in predicting and assessing the course of prostate cancer. Augustus et al., *supra* at 47. Indeed, despite the fact that most prostate cancer patients have carcinomas that are of intermediate grade and stage, prognosis for these types of carcinomas is highly variable. Andrew A Renshaw & Christopher L. Corless, *Prognostic Features in the Pathology of Prostate Cancer, in Prostate Cancer: A Multidisciplinary Guide* 26 (Philip W. Kantoff et al. eds. 1997). Techniques such as transrectal ultrasound, abdominal and pelvic computerized tomography, and MRI have not been particularly useful in predicting local tumor extension. D'Amico, *supra* at 53 (editors' comment). While the use of serum PSA in combination with the Gleason score is currently the most effective method of staging prostate cancer, *id.*, PSA is of limited predictive value, Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26, and the Gleason score is prone to variability and error, King, C. R. & Long, J. P., *Int'l. J. Cancer* 90(6): 326-30 (2000). As such, the current focus of prostate cancer research has been to obtain biomarkers to help better assess the progression of the disease. Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26; Pettaway, C. A., *Tech. Urol.* 4(1): 35-42 (1998).

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop prostate cancer, for diagnosing prostate cancer, for monitoring the progression of the disease, for staging the prostate cancer, for determining whether the prostate cancer has metastasized and for imaging the prostate cancer. There is also a need for better treatment of prostate cancer.

## SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto

that may be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate; identify and monitor prostate tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and  
5 cells, and methods for producing engineered prostate tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to prostate cells and/or prostate tissue. These prostate specific nucleic acids (PSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If  
10 the PSNA is genomic DNA, then the PSNA is a prostate specific gene (PSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 115 through 217. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid  
15 sequence of SEQ ID NO: 1 through 114. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a PSP, or that selectively hybridize or exhibit substantial sequence similarity to a PSNA, as well as allelic variants of a nucleic acid molecule encoding a PSP, and allelic variants of a PSNA. Nucleic acid molecules  
20 comprising a part of a nucleic acid sequence that encodes a PSP or that comprises a part of a nucleic acid sequence of a PSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a PSNA. In a preferred embodiment, the nucleic acid  
25 molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a PSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic  
30 acid molecule encodes all or a fragment of a PSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a PSNA.

-7-

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a PSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a PSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring prostate tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered prostate tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate. The invention provides methods of using the polypeptides of the invention to identify and/or monitor prostate tissue, and to produce engineered prostate tissue.

The agonists and antagonists of the instant invention may be used to treat prostate cancer and non-cancerous disease states in prostate and to produce engineered prostate tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

## 5 DETAILED DESCRIPTION OF THE INVENTION

### Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular  
10 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed  
15 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor  
20 Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4<sup>th</sup> Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of  
25 which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and  
30 techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 5       A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 10       "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 15       The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 20       phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes
- 25       any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.
- 30       A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

10 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

30 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are



-12-

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized  
5 oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate,  
15 phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No.  
20 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is  
25 referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of  
30 RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, e.g., the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, e.g., for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point ( $T_m$ ) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the  $T_m$  for the specific DNA hybrid under a particular set of conditions. The  $T_m$  is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

-15-

The  $T_m$  for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where  $l$  is the length of the hybrid in base pairs.

The  $T_m$  for a particular RNA-RNA hybrid can be estimated by the formula:

5  $T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35$   
(% formamide) - (820/ $l$ ).

The  $T_m$  for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50$$

(% formamide) - (820/ $l$ ).

10 In general, the  $T_m$  decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C  
15 would be subtracted from the calculated  $T_m$  of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without  
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a  
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See Sambrook et al.*

- 5 (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see Sambrook* (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- 15 As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- 20 Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:  $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N)$ , wherein N is change length and the  $[\text{Na}^+]$  is 1 M or less. *See Sambrook* (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the  $T_m$ ) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g., Ausubel* (1999), *supra*; *Sambrook* (1989), *supra*, pp. 11.45-30 11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1  $\mu$ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20  $\mu$ l of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50  $\mu$ g of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a PSP or is a PSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.



"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

5       The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,  
10 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such  
15 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20       The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of  
25 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and  
30 thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that  
5 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding  
10 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in  
15 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-  
25 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a PSP encoded by a nucleic acid molecule of the instant  
30 invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be

5 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single

10 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

15 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared

20 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

25 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such

30 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>-, --CH=CH--(cis and trans), --COCH<sub>2</sub>-, --CH(OH)CH<sub>2</sub>-, and --CH<sub>2</sub>SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2<sup>nd</sup> Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

acids may also be suitable components for polypeptides of the present invention.

Examples of unconventional amino acids include: 4-hydroxyproline,  $\gamma$ -carboxyglutamate,

-N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine,

N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other

- 5 similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

- A protein has "homology" or is "homologous" to a protein from another organism  
10 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"  
15 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence  
20 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

- When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino  
25 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino  
30 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are

5 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is  
15 any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid  
20 substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g., GCG Version 6.1.* Other programs include FASTA, discussed *supra*.

25 A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g., Altschul et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al., Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

30       Expectation value:   10 (default)  
         Filter:               seg (default)  
         Cost to open a gap:  11 (default)



-28-

Cost to extend a gap: 1 (default)  
Max. alignments: 100 (default)  
Word size: 11 (default)  
No. of descriptions: 100 (default)  
5 Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number  
10 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best  
15 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

20 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')<sub>2</sub>, Fv,  
25 dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')<sub>2</sub> fragment is a bivalent fragment comprising two Fab  
30 fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

5 A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

10 The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1  $\mu$ M, preferably less than 100 nM and most preferably less than 10 nM.

15 The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20 The term "prostate specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the prostate as compared to other tissues in the body. In a preferred embodiment, a "prostate specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "prostate specific" nucleic acid molecule or polypeptide is  
25 expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,  
30 such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

*Nucleic Acid Molecules*

5           One aspect of the invention provides isolated nucleic acid molecules that are specific to the prostate or to prostate cells or tissue or that are derived from such nucleic acid molecules. These isolated prostate specific nucleic acids (PSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate, a prostate-specific polypeptide (PSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 115 through 217. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 114.

15           A PSNA may be derived from a human or from another animal. In a preferred embodiment, the PSNA is derived from a human or other mammal. In a more preferred embodiment, the PSNA is derived from a human or other primate. In an even more preferred embodiment, the PSNA is derived from a human.

          By "nucleic acid molecule" for purposes of the present invention, it is also meant  
20   to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a PSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a PSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a PSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively  
25   hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 115 through 217. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 114.

          In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a  
30   nucleic acid molecule encoding a PSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule

-32-

encoding a PSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 115 through 217. In a yet more preferred

5 embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 114. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

10 By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a PSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human PSP. In a more preferred embodiment, the nucleic acid molecule

15 exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 115 through 217. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a PSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 115 through 217, more preferably at least

20 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a PSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid

25 molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a PSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a PSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule

30 comprising a nucleic acid sequence of SEQ ID NO: 1 through 114. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a PSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 114,

more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a PSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a PSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a PSNA or to a nucleic acid molecule encoding a PSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the PSNA or the nucleic acid molecule encoding a PSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 115 through 217 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 114.

The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the PSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a PSNA. Further, the substantially

similar nucleic acid molecule may or may not be a PSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a PSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a PSNA or a nucleic acid encoding a PSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a PSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a PSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 114. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a PSP. However, in a preferred embodiment, the part encodes a PSP. In one aspect, the invention comprises a part of a PSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a PSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a PSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a PSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides.

The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (e.g., reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with  $^{33}\text{P}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ , such as  $^{-32}\text{P}$ -dATP,  $^{-32}\text{P}$ -dCTP,  $^{-32}\text{P}$ -dGTP,  $^{-32}\text{P}$ -dTTP,  $^{-32}\text{P}$ -3'dATP,  $^{-32}\text{P}$ -ATP,  $^{-32}\text{P}$ -CTP,  $^{-32}\text{P}$ -GTP,  $^{-32}\text{P}$ -UTP,  $^{-35}\text{S}$ -dATP,  $\alpha$ - $^{35}\text{S}$ -GTP,  $\alpha$ - $^{33}\text{P}$ -dATP, and the like.



Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301-305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.

Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation,

5 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having  
10 normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019;  
15 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

20 Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a  
25 nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH<sub>2</sub> component parts. Representative U.S. patents  
30 that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;

5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The  $T_m$  of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the  $T_m$  of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the  $T_m$  by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the  $T_m$  by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1):

3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta*. 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure  
5 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24:  
10 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed,  
15 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by  
20 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta*. 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

25

#### *Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic  
30 acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a PSNA, such as deletions, insertions, translocations, and duplications of the PSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g., Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify PSNA in, and isolate PSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A<sup>+</sup>- selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g., Schwarczacher et al., In Situ Hybridization*, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to PSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*;

Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used  
5 as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a PSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 115 through 217.  
10 In another preferred embodiment, the probe or primer is derived from a PSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 114.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17  
15 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using  
20 oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the  
25 art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques,  
30 John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of

which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books  
5 (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion  
10 protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744,  
15 the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by  
20 reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-  
25 planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or  
30 positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another



embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

*Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides*

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous

polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are  
10 incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that  
15 expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or  
20 GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

25 In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial  
30 hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous

derivatives of phage lambda, e.g., NM989,  $\lambda$ GT10 and  $\lambda$ GT11, and other phages, e.g., M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: e.g., typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, e.g. through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp and YEplac series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2,  $2\mu$  plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIp1ac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between

these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally

contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast  $\alpha$ -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the

Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from  
5 the gene comprising the PSNA of interest. Often, expression is enhanced by the incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit  $\beta$ -globin gene and the SV40 splice elements.

10 Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell  
15 genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*,  
20 Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated  
25 by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels  
30 from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline

(Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, e.g., the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, e.g. the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the  $\alpha$ -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, e.g., the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from



those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.* Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)).

Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphiPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide PSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquitination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., [www.expasy.org](http://www.expasy.org) (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

- in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).
- Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.
- In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one  
5 may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website [www.expasy.org](http://www.expasy.org). The nucleic acid molecule is then be introduced into a host cell that is  
10 capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

15 In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the  
20 product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression  
25 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have  
30 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell  
5 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an  
10 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be  
15 able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*  
20 *frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from  
25 *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3  
30 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from prostate are particularly preferred because they may provide a more native  
5 post-translational processing. Particularly preferred are human prostate cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number  
10 of texts and laboratory manuals in the art. *See, e.g.,* Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend  
15 primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.,* Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

20 Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.,* with  $\text{CaCl}_2$ , or a solution of  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Rb}^+$  or  $\text{K}^+$ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent  
25 strains are also available commercially (*e.g.,* Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse  
30 treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols



(BioRad, Richmond, CA, USA) ([http://www.biorad.com/LifeScience/pdf/New\\_Gene\\_Pulser.pdf](http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf)).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the  
5 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and  $\text{Ca}^{2+}$ . Subsequently, the cells are resuspended in a solution of sorbitol, mixed  
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective  
15 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension  
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with  $\text{CaPO}_4$  or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for  $\text{CaPO}_4$  transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated  
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (<http://www.bio-rad.com/LifeScience/pdf/>

- 5 New\_Gene\_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*  
10 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid  
15 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A  
20 Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by  
25 means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

#### Polypeptides

- 30 Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a prostate specific polypeptide (PSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 115 through 217. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a PSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 115 through 217. A polypeptide that comprises only a fragment of an entire PSP may or may not be a polypeptide that is also a PSP. For instance, a full-length polypeptide may be prostate-specific, while a fragment thereof may be found in other tissues as well as in prostate. A polypeptide that is not a PSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-PSP antibodies. However, in a preferred embodiment, the part or fragment is a PSP. Methods of determining whether a polypeptide is a PSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies  
5 (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

10 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75  
15 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a PSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically  
20 synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a  
25 fragment of polypeptide of the invention, preferably a PSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a PSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants,  
30 fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be prostate-specific. In a preferred embodiment, the mutein is prostate-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 115 through 217. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is prostate-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel  
5 (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a PSP. In an even more preferred embodiment, the  
10 polypeptide is homologous to a PSP selected from the group having an amino acid sequence of SEQ ID NO: 115 through 217. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a PSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 115 through 217. In an even  
15 more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%,  
20 more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a PSP comprising an amino acid sequence  
25 of SEQ ID NO: 115 through 217. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a PSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a  
30 PSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSNA is selected from the group consisting of SEQ ID NO: 1 through 114. In another preferred embodiment, the

homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a PSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSP is selected from the group consisting of SEQ ID NO: 115 through 217.

5           The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 115 through 217. The homologous polypeptide may also be a naturally-  
10   occurring polypeptide from a human, when the PSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, e.g., dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring  
15   polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be  
20   one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a PSP. Further, the homologous protein may or may not encode polypeptide that is a PSP. However, in a  
25   preferred embodiment, the homologous polypeptide encodes a polypeptide that is a PSP.

          Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but  
30   also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the

present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a PSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 115 through 217. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 114.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 115 through 217, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);



Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 5       The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS,
- 10   HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC,
- 15   SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- The polypeptides, fragments, and fusion proteins of the present invention can be
- 20   conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 25   usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-PSP antibodies.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 30   usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 115 through 217. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a PSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>--,
- 15 --CH=CH--(cis and trans), --COCH<sub>2</sub>--, --CH(OH)CH<sub>2</sub>-- and --CH<sub>2</sub>SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a PSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000);
- 30 Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- $\beta$ -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

#### *Fusion Proteins*

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 115 through 217, or is a mutein, homologous polypeptide,

analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 114, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 114.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

- of secretion signals and/or leader sequences. For example, a His<sup>6</sup> tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.
- Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbriizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded  
5 protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin  
10 A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins,  $\beta$ -galactosidase, biotin trpE, protein A,  $\beta$ -lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine  
15 at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast  $\phi$ -mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by  
20 enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind  
25 the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the PSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize PSPs, their allelic  
30 variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly PSPs, e.g. by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser



scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of PSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of PSPs.

- 5           One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon  
10 linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-  
15 102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

- Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987).  
20 Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

- Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-  
25 proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

- Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For  
30 example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a PSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 115 through 217, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a PSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a PSP may be indicative of cancer. Differential degradation of the C or N-terminus of a PSP may also be a marker or target for anticancer therapy. For example, a PSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-PSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human prostate.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about  $1 \times 10^{-6}$  molar (M), typically at least about  $5 \times 10^{-7}$  M,  $1 \times 10^{-7}$  M, with affinities and avidities of at least  $1 \times 10^{-8}$  M,  $5 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M and up to  $1 \times 10^{-13}$  M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as  
5 IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such  
10 antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic  
15 mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by  
20 reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered  
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger  
30 mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

25 Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their  
10 entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding  
15 antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B  
20 cells known to be specific for the desired protein, as further described in U.S. Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody  
25 fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein  
30 (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths et al., *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom et al., *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

- 5 Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

- Typically, phage-displayed antibody fragments are scFv fragments or Fab  
10 fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- 15 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.*  
20 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.*  
25 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

- Antibodies and fragments and derivatives thereof of the present invention can  
30 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavalondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

*Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

5       Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

10       Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression  
15       systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57  
20       (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the  
25       proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'<sub>2</sub>, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

30       It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated



nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

5        Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

10        Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature*  
15 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2  
20 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain  
25 diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof,  
30 whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase,  $\beta$ -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide ( $H_2O_2$ ), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

10 As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

20 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,  
25 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,  
30 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as  $^{33}\text{P}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ,  
5 and  $^{125}\text{I}$ .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be  $^{228}\text{Th}$ ,  $^{227}\text{Ac}$ ,  $^{225}\text{Ac}$ ,  $^{223}\text{Ra}$ ,  $^{213}\text{Bi}$ ,  $^{212}\text{Pb}$ ,  $^{212}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{203}\text{Pb}$ ,  $^{194}\text{Os}$ ,  $^{188}\text{Re}$ ,  $^{186}\text{Re}$ ,  $^{153}\text{Sm}$ ,  $^{149}\text{Tb}$ ,  $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{111}\text{In}$ ,  $^{105}\text{Rh}$ ,  $^{99\text{m}}\text{Tc}$ ,  $^{97}\text{Ru}$ ,  $^{90}\text{Y}$ ,  $^{90}\text{Sr}$ ,  $^{88}\text{Y}$ ,  $^{72}\text{Se}$ ,  $^{67}\text{Cu}$ , or  $^{47}\text{Sc}$ .

10 As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the  
15 application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria*  
20 toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate,  
25 and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more  
30 of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to  
5 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in  
10 prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind  
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present  
20 invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody  
25 molecule, or to alter it in any other way that may render it more suitable for a particular application.

#### Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human  
30 organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a PSP. In a preferred embodiment, the PSP comprises an amino acid

sequence selected from SEQ ID NO: 115 through 217, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a PSNA of the invention, preferably a PSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 114, or a part, substantially similar nucleic acid molecule, allelic variant  
5 or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human PSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human  
10 organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A  
15 Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent  
20 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of  
30 nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campbell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene  
5 may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

10 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using  
15 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

20 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels  
25 because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is  
30 appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC



compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt  
5 the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

10 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

15 Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

20 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the  
25 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such  
30 conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 114 and SEQ ID NO: 115 through 217 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

#### Diagnostic Methods for Prostate Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of a PSNA or a PSP in a human patient that has or may have prostate cancer, or who is at risk of developing prostate cancer, with the expression of a PSNA or a PSP in a normal human control. For purposes of the present invention, "expression of a PSNA" or "PSNA expression" means the quantity of PSG mRNA that  
5 can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a PSP" or "PSP expression" means the amount of PSP that can be measured by any method known in the art or the level of translation of a PSG PSNA that can be measured by any method known in the art.

10 The present invention provides methods for diagnosing prostate cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of PSNA or PSP in cells, tissues, organs or bodily fluids compared with levels of PSNA or PSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a PSNA or  
15 PSP in the patient versus the normal human control is associated with the presence of prostate cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in the structure of the mRNA of a PSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations  
20 in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in a PSP compared to a PSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the PSP or subcellular PSP localization.

25 In a preferred embodiment, the expression of a PSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 115 through 217, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the PSNA expression that is measured is the level of expression of a PSNA mRNA selected from SEQ ID NO: 1 through 114, or a hybridizing nucleic  
30 acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. PSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative

or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. PSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a PSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, PSNA expression may be compared to a known control, such as normal prostate nucleic acid, to detect a change in expression.

10 In another preferred embodiment, the expression of a PSP is measured by determining the level of a PSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 115 through 217, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of PSNA or PSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of prostate cancer. The expression level of a PSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the PSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.*, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the PSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

25 In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a PSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-PSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the PSP will bind to the anti-PSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-PSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the PSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a PSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure PSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-PSP antibody is attached to a solid support and an allocated amount of a labeled PSP and a sample of interest are incubated with the solid support. The amount of labeled PSP detected which is attached to the solid support can be correlated to the quantity of a PSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a PSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more PSNAs of interest. In this approach, all or a portion of one or more PSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of PSNA or PSP includes, without limitation, prostate tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, prostate cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary prostate cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See Scott, supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in PSNAs or PSPs in cells in sputum samples may be particularly

useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a PSNA or PSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other PSNA or PSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular PSNA or PSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

#### *Diagnosing*

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having prostate cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a PSNA and/or PSP and then ascertaining whether the patient has prostate cancer from the expression level of the PSNA or PSP. In general, if high expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether prostate cancer has metastasized in a patient. One may identify whether the prostate cancer has



metastasized by measuring the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a variety of tissues. The presence of a PSNA or PSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a PSNA or PSP is associated with prostate cancer. Similarly, the presence of a PSNA or PSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a PSNA or PSP is associated with prostate cancer. Further, the presence of a structurally altered PSNA or PSP that is associated with prostate cancer is also indicative of metastasis.

10 In general, if high expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

20 The PSNA or PSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with prostate cancers or other prostate related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of prostate disorders.

25

### *Staging*

The invention also provides a method of staging prostate cancer in a human patient. The method comprises identifying a human patient having prostate cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more PSNAs or PSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more PSNAs or PSPs is determined for each stage to

-101-

obtain a standard expression level for each PSNA and PSP. Then, the PSNA or PSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The PSNA or PSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the PSNAs and PSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a PSNA or PSP to determine the stage of a prostate cancer.

### *Monitoring*

Further provided is a method of monitoring prostate cancer in a human patient.

One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the prostate cancer. The method comprises identifying a human patient that one wants to monitor for prostate cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more PSNAs or PSPs, and comparing the PSNA or PSP levels over time to those PSNA or PSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a PSNA or PSP that are associated with prostate cancer.

If increased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of PSNAs or PSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

Monitoring a patient for onset of prostate cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a PSNA and/or PSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more PSNAs and/or PSPs are detected. The presence of higher (or lower) PSNA or PSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly prostate cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more PSNAs and/or PSPs of the invention can also be monitored by analyzing levels of expression of the PSNAs and/or PSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

#### *Detection of Genetic Lesions or Mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in a PSG, thereby determining if a human with the genetic lesion is susceptible to developing prostate cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing prostate cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the PSGs of this invention, a chromosomal rearrangement of PSG, an aberrant modification of PSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a PSG. Methods to detect such lesions in the PSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

#### Methods of Detecting Noncancerous Prostate Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having or known to have a noncancerous prostate disease. In general, the method comprises the steps of obtaining a sample from the patient,

determining the expression level or structural alterations of a PSNA and/or PSP, comparing the expression level or structural alteration of the PSNA or PSP to a normal prostate control, and then ascertaining whether the patient has a noncancerous prostate disease. In general, if high expression relative to a control of a PSNA or PSP is  
5 indicative of a particular noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is  
10 indicative of a noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same  
15 patient.

One having ordinary skill in the art may determine whether a PSNA and/or PSP is associated with a particular noncancerous prostate disease by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining which PSNAs and/or PSPs are expressed in the tissue at either a higher or a lower level than in  
20 normal prostate tissue. In another embodiment, one may determine whether a PSNA or PSP exhibits structural alterations in a particular noncancerous prostate disease state by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining the structural alterations in one or more PSNAs and/or PSPs relative to normal prostate tissue.

25 Methods for Identifying Prostate Tissue

In another aspect, the invention provides methods for identifying prostate tissue. These methods are particularly useful in, e.g., forensic science, prostate cell  
30 differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is prostate tissue or has prostate tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising prostate tissue or

- having prostate tissue-like characteristics, determining whether the sample expresses one or more PSNAs and/or PSPs, and, if the sample expresses one or more PSNAs and/or PSPs, concluding that the sample comprises prostate tissue. In a preferred embodiment, the PSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID
- 5 NO: 115 through 217, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 114, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a PSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot
- 10 hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a PSP is expressed. Determining whether a sample expresses a PSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the PSP has an amino acid sequence selected from SEQ ID NO: 115
- 15 through 217, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two PSNAs and/or PSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five PSNAs and/or PSPs are determined.

- In one embodiment, the method can be used to determine whether an unknown
- 20 tissue is prostate tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into prostate tissue. This is important in monitoring the effects of the addition of various agents to cell or
- 25 tissue culture, *e.g.*, in producing new prostate tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Prostate Tissue

In another aspect, the invention provides methods for producing engineered prostate tissue or cells. In one embodiment, the method comprises the steps of providing  
5 cells, introducing a PSNA or a PSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of prostate tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal prostate tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered prostate tissue or cells comprises one of these cell types. In another  
10 embodiment, the engineered prostate tissue or cells comprises more than one prostate cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the prostate cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more PSPs are introduced into cells,  
15 preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode PSPs having amino acid sequences selected from SEQ ID NO: 115 through 217, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 114, or hybridizing nucleic acids, allelic variants or parts  
20 thereof. In another highly preferred embodiment, a PSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial prostate tissue may be used to treat patients who have lost some or all of their prostate function.

25 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a  
30 preferred embodiment, the pharmaceutical composition comprises a PSNA or part thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 114, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity

thereto. In another preferred embodiment, the pharmaceutical composition comprises a PSP or fragment thereof. In a more preferred embodiment, the PSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 115 through 217, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-PSP antibody, preferably an antibody that specifically binds to a PSP having an amino acid that is selected from the group consisting of SEQ ID NO: 115 through 217, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20<sup>th</sup> ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7<sup>th</sup> ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3<sup>rd</sup> ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium

carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

- 5 Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

- Tablet binders that can be used include acacia, methylcellulose, sodium  
10 carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

- Fillers, agents that facilitate disintegration and/or solubilization, tablet binders  
15 and lubricants, including the aforementioned, can be used singly or in combination.

- Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic  
20 solvents or solvent mixtures.

- Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and,  
25 optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

- 30 Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin,



carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

- 5           The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 10 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

- Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a 15 pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as 20 sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

- 25           Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

- 30           Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as

polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that  
5 are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and  
10 can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin  
15 penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases  
20 as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various  
25 powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for  
30 reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding  
5 free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of  
10 those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example PSP polypeptide, fusion protein, or fragments thereof, antibodies specific for PSP, agonists, antagonists or inhibitors of PSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical  
15 arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of  
20 administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical  
25 compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of  
30 active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

#### Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, e.g., in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of prostate function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

*Gene Therapy and Vaccines*

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. See, e.g., Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a PSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a PSP are administered, for example, to complement a deficiency in the native PSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. See, e.g., Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a PSP having the amino acid sequence of SEQ ID NO: 115 through 217, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a PSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in PSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a PSP having the amino acid sequence of SEQ ID NO: 115 through 217, or a fragment, fusion protein, allelic variant or homolog thereof.

### *Antisense Administration*

Antisense nucleic acid compositions, or vectors that drive expression of a PSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a PSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a PSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to PSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the PSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a PSP, preferably a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 114, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

### *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a PSP, a

fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant PSP defect.

Protein compositions are administered, for example, to complement a deficiency in native PSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to PSP. The immune response can be used to modulate activity of PSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate PSP.

10 In a preferred embodiment, the polypeptide is a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 114, or a part, allelic variant, substantially similar or hybridizing nucleic acid  
15 thereof.

#### *Antibody, Agonist and Antagonist Administration*

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is  
20 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of PSP, or to target therapeutic agents to sites of PSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred  
25 embodiment, the antibody specifically binds to a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 114, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a PSP or have a modulatory effect on the expression or activity of a PSP.  
30 Modulators which decrease the expression or activity of PSP (antagonists) are believed to be useful in treating prostate cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small

molecules predicted via computer imaging to specifically bind to regions of a PSP can also be designed, synthesized and tested for use in the imaging and treatment of prostate cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the PSPs identified herein. Molecules  
5 identified in the library as being capable of binding to a PSP are key candidates for further evaluation for use in the treatment of prostate cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a PSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of PSP is  
10 administered. Antagonists of PSP can be produced using methods generally known in the art. In particular, purified PSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a PSP.

In other embodiments a pharmaceutical composition comprising an agonist of a  
15 PSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP comprising an amino acid sequence of SEQ ID NO: 115 through 217, or a fusion protein, allelic variant, homolog, analog or  
20 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 114, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

#### *Targeting Prostate Tissue*

25 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the prostate or to specific cells in the prostate. In a preferred embodiment, an anti-PSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if prostate tissue needs to be  
30 selectively destroyed. This would be useful for targeting and killing prostate cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting prostate cell function.



In another embodiment, an anti-PSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring prostate function, identifying prostate cancer tumors, and identifying noncancerous prostate diseases.

5

## EXAMPLES

### Example 1: Gene Expression analysis

PSGs were identified by mRNA subtraction analysis using standard methods.

The sequences were extended using GeneBank sequences, Incyte's proprietary database.

From the nucleotide sequences, predicted amino acid sequences were prepared.

- 10 DEX0281\_1, DEX0281\_2 correspond to SEQ ID NO.1, 2 etc. DEX0127 was the parent sequence found in the mRNA subtractions.

	DEX0283_1	DEX0128_1	DEX0283_115
	DEX0283_2	DEX0128_2	DEX0283_116
	DEX0283_3	DEX0128_3	DEX0283_117
15	DEX0283_4	DEX0128_4	DEX0283_118
	DEX0283_5	flex DEX0128_4	DEX0283_119
	DEX0283_6	DEX0128_5	DEX0283_120
	DEX0283_7	DEX0128_6	DEX0283_121
	DEX0283_8	DEX0128_7	DEX0283_122
20	DEX0283_9	DEX0128_8	DEX0283_123
	DEX0283_10	flex DEX0128_8	
	DEX0283_11	DEX0128_9	DEX0283_124
	DEX0283_12	DEX0128_10	DEX0283_125
	DEX0283_13	DEX0128_11	DEX0283_126
25	DEX0283_14	DEX0128_12	DEX0283_127
	DEX0283_15	DEX0128_13	DEX0283_128
	DEX0283_16	DEX0128_14	DEX0283_129
	DEX0283_17	DEX0128_15	DEX0283_130
	DEX0283_18	DEX0128_16	DEX0283_131
30	DEX0283_19	flex DEX0128_16	DEX0283_132
	DEX0283_20	DEX0128_17	DEX0283_133
	DEX0283_21	flex DEX0128_17	DEX0283_134
	DEX0283_22	DEX0128_18	DEX0283_135
	DEX0283_23	DEX0128_19	DEX0283_136
35	DEX0283_24	DEX0128_20	DEX0283_137
	DEX0283_25	DEX0128_21	DEX0283_138
	DEX0283_26	DEX0128_22	DEX0283_139
	DEX0283_27	DEX0128_23	DEX0283_140
	DEX0283_28	DEX0128_24	DEX0283_141
40	DEX0283_29	DEX0128_25	DEX0283_142
	DEX0283_30	flex DEX0128_25	DEX0283_143

	DEX0283_31	DEX0128_26 DEX0283_144
	DEX0283_32	DEX0128_27 DEX0283_145
	DEX0283_33	DEX0128_28 DEX0283_146
	DEX0283_34	DEX0128_29 DEX0283_147
5	DEX0283_35	DEX0128_30 DEX0283_148
	DEX0283_36	flex DEX0128_30 DEX0283_149
	DEX0283_37	DEX0128_31 DEX0283_150
	DEX0283_38	DEX0128_32 DEX0283_151
	DEX0283_39	DEX0128_33 DEX0283_152
10	DEX0283_40	DEX0128_34 DEX0283_153
	DEX0283_41	DEX0128_35 DEX0283_154
	DEX0283_42	DEX0128_36 DEX0283_155
	DEX0283_43	flex DEX0128_36 DEX0283_156
	DEX0283_44	DEX0128_37 DEX0283_157
15	DEX0283_45	DEX0128_38 DEX0283_158
	DEX0283_46	DEX0128_39 DEX0283_159
	DEX0283_47	DEX0128_40 DEX0283_160
	DEX0283_48	DEX0128_41 DEX0283_161
	DEX0283_49	DEX0128_42 DEX0283_162
20	DEX0283_50	DEX0128_43 DEX0283_163
	DEX0283_51	DEX0128_44 DEX0283_164
	DEX0283_52	DEX0128_45 DEX0283_165
	DEX0283_53	DEX0128_46 DEX0283_166
	DEX0283_54	DEX0128_47 DEX0283_167
25	DEX0283_55	flex DEX0128_47
	DEX0283_56	DEX0128_48 DEX0283_168
	DEX0283_57	DEX0128_49 DEX0283_169
	DEX0283_58	DEX0128_50 DEX0283_170
	DEX0283_59	DEX0128_51 DEX0283_171
30	DEX0283_60	flex DEX0128_51
	DEX0283_61	DEX0128_52 DEX0283_172
	DEX0283_62	DEX0128_53 DEX0283_173
	DEX0283_63	DEX0128_54 DEX0283_174
	DEX0283_64	DEX0128_55 DEX0283_175
35	DEX0283_65	DEX0128_56 DEX0283_176
	DEX0283_66	DEX0128_57 DEX0283_177
	DEX0283_67	DEX0128_58 DEX0283_178
	DEX0283_68	flex DEX0128_58 DEX0283_179
	DEX0283_69	DEX0128_59 DEX0283_180
40	DEX0283_70	DEX0128_60 DEX0283_181
	DEX0283_71	DEX0128_61 DEX0283_182
	DEX0283_72	DEX0128_62 DEX0283_183
	DEX0283_73	DEX0128_63 DEX0283_184
	DEX0283_74	DEX0128_64 DEX0283_185
45	DEX0283_75	DEX0128_65 DEX0283_186
	DEX0283_76	DEX0128_66 DEX0283_187
	DEX0283_77	DEX0128_67 DEX0283_188
	DEX0283_78	DEX0128_68 DEX0283_189

-118-

	DEX0283_79	DEX0128_69	DEX0283_190
	DEX0283_80	DEX0128_70	DEX0283_191
	DEX0283_81	DEX0128_71	DEX0283_192
	DEX0283_82	DEX0128_72	DEX0283_193
5	DEX0283_83	DEX0128_73	DEX0283_194
	DEX0283_84	DEX0128_74	DEX0283_195
	DEX0283_85	DEX0128_75	
	DEX0283_86	DEX0128_76	
	DEX0283_87	DEX0128_77	DEX0283_196
10	DEX0283_88	DEX0128_78	DEX0283_197
	DEX0283_89	DEX0128_79	DEX0283_198
	DEX0283_90	DEX0128_80	DEX0283_199
	DEX0283_91	DEX0128_81	
	DEX0283_92	DEX0128_82	
15	DEX0283_93	DEX0128_83	DEX0283_200
	DEX0283_94	DEX0128_84	DEX0283_201
	DEX0283_95	DEX0128_85	DEX0283_202
	DEX0283_96	DEX0128_86	DEX0283_203
	DEX0283_97	DEX0128_87	DEX0283_204
20	DEX0283_98	DEX0128_88	DEX0283_205
	DEX0283_99	DEX0128_89	DEX0283_206
	DEX0283_100	DEX0128_90	DEX0283_207
	DEX0283_101	flex DEX0128_90	
	DEX0283_102	DEX0128_91	DEX0283_208
25	DEX0283_103	flex DEX0128_91	
	DEX0283_104	DEX0128_92	DEX0283_209
	DEX0283_105	DEX0128_93	DEX0283_210
	DEX0283_106	DEX0128_94	DEX0283_211
	DEX0283_107	DEX0128_95	DEX0283_212
30	DEX0283_108	DEX0128_96	DEX0283_213
	DEX0283_109	flex DEX0128_96	
	DEX0283_110	DEX0128_97	DEX0283_214
	DEX0283_111	DEX0128_98	DEX0283_215
	DEX0283_112	DEX0128_99	DEX0283_216
35	DEX0283_113	flex DEX0128_99	
	DEX0283_114	DEX0128_100	DEX0283_217

The predicted chromosomal locations are as follows:

	DEX0283_3	chromosome 10
40	DEX0283_4	chromosome 9
	DEX0283_5	chromosome X
	DEX0283_7	chromosome 6
	DEX0283_8	chromosome 3
	DEX0283_9	chromosome 20
45	DEX0283_10	chromosome 20
	DEX0283_11	chromosome 2
	DEX0283_13	chromosome 8
	DEX0283_14	chromosome 8

-119-

DEX0283\_16 chromosome 9  
 DEX0283\_19 chromosome 2  
 DEX0283\_25 chromosome 8  
 DEX0283\_29 chromosome 18  
 5 DEX0283\_30 chromosome 18  
 DEX0283\_33 chromosome 11  
 DEX0283\_39 chromosome 7  
 DEX0283\_42 chromosome 1  
 DEX0283\_43 chromosome 2  
 10 DEX0283\_45 chromosome 11  
 DEX0283\_46 chromosome 14  
 DEX0283\_51 chromosome 9  
 DEX0283\_53 chromosome 10  
 DEX0283\_54 chromosome 5  
 15 DEX0283\_55 chromosome 5  
 DEX0283\_57 chromosome 5  
 DEX0283\_60 chromosome 3  
 DEX0283\_64 chromosome 4  
 DEX0283\_65 chromosome 9  
 20 DEX0283\_67 chromosome 1  
 DEX0283\_68 chromosome 2  
 DEX0283\_69 chromosome 13  
 DEX0283\_75 chromosome 4  
 DEX0283\_77 chromosome X  
 25 DEX0283\_78 chromosome 8  
 DEX0283\_80 chromosome 9  
 DEX0283\_83 chromosome 1  
 DEX0283\_84 chromosome 1  
 DEX0283\_85 chromosome 19  
 30 DEX0283\_92 chromosome 8  
 DEX0283\_93 chromosome 15  
 DEX0283\_101 chromosome 4  
 DEX0283\_102 chromosome 6  
 DEX0283\_103 chromosome 6  
 35 DEX0283\_104 chromosome 7  
 DEX0283\_108 chromosome 8  
 DEX0283\_109 chromosome 8  
 DEX0283\_113 chromosome 7

#### 40 **Example 2: Relative Quantitation of Gene Expression**

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation  
 detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The  
 method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5'  
 reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity  
 45 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected

by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this  
5 endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence  
10 Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain  
15 reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of  
20 expression of the PSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the PSNA in pairs of matching samples and 1  
25 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the PSNAs that show a high degree of tissue  
30 specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

-121-

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

- 5 Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 114 being a diagnostic marker for cancer.

### Example 3: Protein Expression

- 10 The PSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the PSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the PSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH<sub>2</sub>-terminus of the coding sequence of PSNA, and six histidines, flanking the COOH-terminus of the coding sequence of PSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

- 15 An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

- Large-scale purification of PSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. PSP was eluted stepwise with various concentration imidazole buffers.
- 20

### Example 4: Protein Fusions

- 25 Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a
- 30

polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second  
5 signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

#### **Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse)  
10 with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The  
15 splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*,  
20 *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes  
25 use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability  
30 to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

5	DEX0283_116	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	13-27	1.44	15	
	DEX0283_124	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	30-46	1.20	17	
10	DEX0283_125	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	33-56	1.05	24	
	DEX0283_127	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	2-21	1.21	20	
15	DEX0283_129	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	22-51	1.05	30	
20	DEX0283_132	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	103-113	1.13	11	
	73-84	1.05	12	
	164-182	1.02	19	
	DEX0283_144	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	7-16	1.19	10	
25	DEX0283_149	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	91-103	1.14	13	
30	32-84	1.09	53	
	DEX0283_150	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	4-14	1.13	11	
	DEX0283_152	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	5-25	1.01	21	
35	DEX0283_154	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	6-24	1.09	19	
40	DEX0283_159	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	44-55	1.03	12	
	DEX0283_161	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	12-21	1.09	10	
45	DEX0283_163	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	39-51	1.11	13	
	DEX0283_164	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	7-24	1.15	18	
50	DEX0283_165	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	10-35	1.01	26	
55	DEX0283_169	Antigenicity Index(Jameson-Wolf)		



-124-

		positions	AI avg	length
		15-36	1.20	22
	DEX0283_170	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
5		79-99	1.26	21
	DEX0283_174	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		13-22	1.32	10
	DEX0283_175	Antigenicity Index(Jameson-Wolf)		
10		positions	AI avg	length
		25-36	1.09	12
	DEX0283_179	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		295-305	1.19	11
15		168-194	1.14	27
		35-50	1.14	16
		65-85	1.12	21
		448-494	1.02	47
		428-443	1.01	16
20	DEX0283_187	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		24-38	1.03	15
	DEX0283_190	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
25		43-69	1.12	27
	DEX0283_195	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		13-26	1.05	14
	DEX0283_196	Antigenicity Index(Jameson-Wolf)		
30		positions	AI avg	length
		11-20	1.46	10
		58-67	1.13	10
	DEX0283_207	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
35		41-51	1.11	11
	DEX0283_208	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		7-28	1.04	22
	DEX0283_211	Antigenicity Index(Jameson-Wolf)		
40		positions	AI avg	length
		32-50	1.03	19
	DEX0283_212	Antigenicity Index(Jameson-Wolf)		
		positions	AI avg	length
		36-49	1.05	14
45		3-34	1.00	32

Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the

50 ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs of the invention ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_prosite.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html))

most recently accessed October 23, 2001). For full definitions of the PTMs see  
<http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

- DEX0283\_115 Asn\_Glycosylation 14-17;
- DEX0283\_116 Asn\_Glycosylation 55-58; Camp\_Phospho\_Site 40-43; Ck2\_Phospho\_Site 39-42;43-46;
- 5 DEX0283\_117 Pkc\_Phospho\_Site 17-19;39-41;
- DEX0283\_118 Myristyl 32-37; Pkc\_Phospho\_Site 6-8;9-11;20-22; Tyr\_Phospho\_Site 22-29;
- DEX0283\_119 Myristyl 21-26; Pkc\_Phospho\_Site 7-9;
- DEX0283\_119 Ck2\_Phospho\_Site 13-16; Leucine\_Zipper 22-43; Myristyl 10-15; Pkc\_Phospho\_Site 13-15;
- 10 DEX0283\_120 Asn\_Glycosylation 24-27; Ck2\_Phospho\_Site 43-46;
- DEX0283\_122 Asn\_Glycosylation 36-39; Ck2\_Phospho\_Site 44-47; Pkc\_Phospho\_Site 38-40;
- Tyr\_Phospho\_Site 33-41;
- DEX0283\_123 Pkc\_Phospho\_Site 22-24;
- DEX0283\_125 Pkc\_Phospho\_Site 54-56; Tyr\_Phospho\_Site 62-70;
- 15 DEX0283\_126 Myristyl 9-14; Pkc\_Phospho\_Site 27-29;
- DEX0283\_127 Asn\_Glycosylation 11-14;14-17;
- DEX0283\_128 Amidation 26-29;
- DEX0283\_129 Ck2\_Phospho\_Site 3-6; Pkc\_Phospho\_Site 3-5;10-12;23-25;51-53;
- DEX0283\_130 Tyr\_Phospho\_Site 14-21;
- 20 DEX0283\_132 Amidation 111-114; Ck2\_Phospho\_Site 29-32;61-64;63-66;127-130;227-230; Myristyl 199-204;293-298;297-302; Pkc\_Phospho\_Site 55-57;74-76;210-212;214-216;227-229;248-250;279-281; Tyr\_Phospho\_Site 299-306;
- DEX0283\_133 Ck2\_Phospho\_Site 22-25; Pkc\_Phospho\_Site 22-24;
- DEX0283\_134 Asn\_Glycosylation 12-15; Ck2\_Phospho\_Site 31-34;49-52; Pkc\_Phospho\_Site 18-20;21-23;53-55;58-60;66-68;
- 25 DEX0283\_135 Myristyl 14-19;
- DEX0283\_136 Asn\_Glycosylation 31-34;
- DEX0283\_138 Amidation 9-12; Asn\_Glycosylation 19-22;37-40; Pkc\_Phospho\_Site 13-15;
- DEX0283\_139 Ck2\_Phospho\_Site 58-61;63-66; Myristyl 2-7;20-25;
- 30 DEX0283\_140 Pkc\_Phospho\_Site 41-43;
- DEX0283\_141 Pkc\_Phospho\_Site 10-12;
- DEX0283\_142 Pkc\_Phospho\_Site 9-11;
- DEX0283\_143 Myristyl 18-23; Prokar\_Lipoprotein 20-30;
- DEX0283\_145 Amidation 26-29;
- 35 DEX0283\_146 Pkc\_Phospho\_Site 18-20;
- DEX0283\_147 Myristyl 8-13; Pkc\_Phospho\_Site 12-14;
- DEX0283\_148 Amidation 58-61; Asn\_Glycosylation 22-25; Myristyl 58-63; Pkc\_Phospho\_Site 11-13;42-44;
- DEX0283\_149 Ck2\_Phospho\_Site 20-23;58-61; Myristyl 53-58; Pkc\_Phospho\_Site 57-59;
- 40 DEX0283\_150 Asn\_Glycosylation 10-13;11-14; Myristyl 30-35; Pkc\_Phospho\_Site 20-22;31-33;
- DEX0283\_152 Asn\_Glycosylation 13-16; Pkc\_Phospho\_Site 5-7;
- DEX0283\_153 Myristyl 2-7;
- DEX0283\_154 Asn\_Glycosylation 6-9; Pkc\_Phospho\_Site 15-17;
- DEX0283\_155 Asn\_Glycosylation 27-30;44-47; Myristyl 39-44; Pkc\_Phospho\_Site 2-4;8-10;29-31;
- 45 DEX0283\_156 Ig\_Mhc 9-15; Myristyl 62-67;
- DEX0283\_157 Camp\_Phospho\_Site 49-52; Pkc\_Phospho\_Site 47-49;52-54;66-68;
- DEX0283\_158 Asn\_Glycosylation 26-29; Pkc\_Phospho\_Site 9-11;
- DEX0283\_159 Asn\_Glycosylation 3-6; Camp\_Phospho\_Site 29-32; Ck2\_Phospho\_Site 44-47;
- Pkc\_Phospho\_Site 27-29;34-36;
- 50 DEX0283\_160 Asn\_Glycosylation 26-29;
- DEX0283\_163 Myristyl 59-64;
- DEX0283\_164 Pkc\_Phospho\_Site 16-18;
- DEX0283\_165 Amidation 58-61; Asn\_Glycosylation 47-50; Myristyl 24-29; Pkc\_Phospho\_Site 32-34;35-37;
- 55 DEX0283\_166 Myristyl 11-16;
- DEX0283\_167 Pkc\_Phospho\_Site 3-5;

- DEX0283\_168 Pkc\_Phospho\_Site 31-33;  
 DEX0283\_169 Pkc\_Phospho\_Site 31-33;  
 DEX0283\_170 Camp\_Phospho\_Site 95-98; Pkc\_Phospho\_Site 94-96;104-106;  
 DEX0283\_171 Ck2\_Phospho\_Site 17-20; Myristyl 35-40;  
 5 DEX0283\_172 Myristyl 67-72;70-75; Pkc\_Phospho\_Site 41-43;  
 DEX0283\_173 Ck2\_Phospho\_Site 16-19; Tyr\_Phospho\_Site 5-12;  
 DEX0283\_174 Ck2\_Phospho\_Site 14-17; Myristyl 39-44; Pkc\_Phospho\_Site 6-8;  
 DEX0283\_175 Asn\_Glycosylation 20-23;28-31; Myristyl 11-16;34-39; Pkc\_Phospho\_Site 30-32;51-53;  
 DEX0283\_176 Asn\_Glycosylation 10-13; Myristyl 22-27; Pkc\_Phospho\_Site 31-33;  
 10 DEX0283\_177 Pkc\_Phospho\_Site 46-48;  
 DEX0283\_178 Asn\_Glycosylation 57-60; Ck2\_Phospho\_Site 47-50; Myristyl 70-75; Pkc\_Phospho\_Site 59-61; Tyr\_Phospho\_Site 80-87;  
 DEX0283\_179 Amidation 175-178;296-299; Ck2\_Phospho\_Site 104-107;160-163; Glycosaminoglycan 74-77;302-305; Myristyl 71-76;75-80;76-81;199-204;245-250;510-515;  
 15 Pkc\_Phospho\_Site 14-16;80-82;81-83;210-212;221-223;378-380;492-494;526-528;  
 Tyr\_Phospho\_Site 468-474;  
 DEX0283\_180 Ck2\_Phospho\_Site 43-46; Pkc\_Phospho\_Site 10-12;30-32;43-45;  
 DEX0283\_181 Pkc\_Phospho\_Site 36-38;  
 DEX0283\_182 Asn\_Glycosylation 23-26;37-40; Pkc\_Phospho\_Site 48-50;  
 20 DEX0283\_184 Ck2\_Phospho\_Site 84-87; Myristyl 29-34;37-42;66-71; Prokar\_Lipoprotein 99-109;  
 DEX0283\_185 Ck2\_Phospho\_Site 2-5; Myristyl 29-34;37-42;  
 DEX0283\_186 Asn\_Glycosylation 2-5; Ck2\_Phospho\_Site 3-6; Pkc\_Phospho\_Site 21-23;  
 DEX0283\_187 Ck2\_Phospho\_Site 40-43; Pkc\_Phospho\_Site 30-32;  
 DEX0283\_188 Asn\_Glycosylation 49-52;60-63; Pkc\_Phospho\_Site 14-16;25-27;51-53;  
 25 DEX0283\_189 Asn\_Glycosylation 17-20;  
 DEX0283\_190 Pkc\_Phospho\_Site 22-24;110-112;  
 DEX0283\_193 Asn\_Glycosylation 26-29; Myristyl 6-11; Pkc\_Phospho\_Site 29-31;  
 DEX0283\_194 Myristyl 37-42;  
 DEX0283\_195 Camp\_Phospho\_Site 17-20;  
 30 DEX0283\_196 Asn\_Glycosylation 52-55; Pkc\_Phospho\_Site 30-32;64-66;  
 DEX0283\_197 Myristyl 27-32;  
 DEX0283\_198 Pkc\_Phospho\_Site 3-5;  
 DEX0283\_200 Amidation 96-99; Ck2\_Phospho\_Site 11-14;92-95; Glycosaminoglycan 76-79; Myristyl 2-7;96-101;100-105; Pkc\_Phospho\_Site 7-9;27-29;44-46;55-57;80-82;  
 35 DEX0283\_201 Ck2\_Phospho\_Site 45-48; Pkc\_Phospho\_Site 35-37; Prokar\_Lipoprotein 15-25;  
 DEX0283\_202 Ck2\_Phospho\_Site 46-49;48-51; Pkc\_Phospho\_Site 15-17;26-28;72-74;  
 DEX0283\_203 Asn\_Glycosylation 36-39;46-49; Camp\_Phospho\_Site 22-25; Ck2\_Phospho\_Site 8-11;  
 Pkc\_Phospho\_Site 8-10;20-22;48-50;  
 DEX0283\_204 Amidation 25-28; Ck2\_Phospho\_Site 5-8;  
 40 DEX0283\_205 Ck2\_Phospho\_Site 6-9;24-27;  
 DEX0283\_206 Pkc\_Phospho\_Site 16-18;  
 DEX0283\_207 Asn\_Glycosylation 50-53; Myristyl 48-53; Pkc\_Phospho\_Site 40-42;  
 DEX0283\_208 Asn\_Glycosylation 13-16; Pkc\_Phospho\_Site 16-18;  
 DEX0283\_209 Myristyl 48-53; Pkc\_Phospho\_Site 2-4;  
 45 DEX0283\_210 Asn\_Glycosylation 36-39; Pkc\_Phospho\_Site 4-6;40-42;  
 DEX0283\_211 Asn\_Glycosylation 48-51; Pkc\_Phospho\_Site 32-34;  
 DEX0283\_212 Camp\_Phospho\_Site 5-8; Myristyl 57-62;72-77;92-97; Pkc\_Phospho\_Site 61-63;  
 DEX0283\_213 Asn\_Glycosylation 5-8; Glycosaminoglycan 7-10;  
 DEX0283\_214 Camp\_Phospho\_Site 11-14; Pkc\_Phospho\_Site 14-16;  
 50 DEX0283\_217 Asn\_Glycosylation 12-15; Ck2\_Phospho\_Site 7-10;52-55; Myristyl 74-79;  
 Pkc\_Phospho\_Site 35-37;

**Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 114. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and

translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

- 5           Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The
- 10       coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature.
- 15       The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

- The reaction is measured by a microtiter plate reader. A standard curve is
- 20       prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

**Example 8: Formulating a Polypeptide**

- 25       The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for
- 30       purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 ,  $\mu\text{g/kg/day}$  to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and

EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

5 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

10 For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more  
15 preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that  
20 enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic  
25 polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

30 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about

3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### **Example 9: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100  $\mu\text{g/kg}$  of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.



**Example 10: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

- 5 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

10 **Example 11: Method of Treatment Using Gene Therapy**

- One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned  
15 upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one  
20 week.

- At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine  
25 sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

- The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set  
30 forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the

presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

5           The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are  
10 now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is  
15 removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently  
20 infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

#### **Example 12: Method of Treatment Using Gene Therapy-*In Vivo***

Another aspect of the present invention is using *in vivo* gene therapy methods to  
25 treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide  
30 by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et

al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers  
5 injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell,  
10 including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

15 The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target  
20 cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow,  
25 thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue  
30 ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They

-135-

may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

### 10 Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campbell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810-813 (1997)).

5       The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a  
10       particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous  
15       gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus  
20       inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant  
25       gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples  
30       obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated

immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### **Example 14: Knock-Out Animals**

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications

to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

5 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including  
10 human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence  
15 associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the  
20 control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the  
25 body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

30 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the



-140-

cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include,  
5 but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein  
10 are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The  
15 present invention is limited only by the claims that follow.

## CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
  - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes  
5 an amino acid sequence of SEQ ID NO: 115 through 217;
  - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID  
NO: 1 through 114;
  - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid  
molecule of (a) or (b); or
  - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic  
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is a cDNA.  
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid  
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid  
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a prostate specific nucleic acid  
(PSNA) in a sample, comprising the steps of:
  - (a) contacting the sample with the nucleic acid molecule according to claim 1  
under conditions in which the nucleic acid molecule will selectively hybridize to a  
prostate specific nucleic acid; and
  - 30 (b) detecting hybridization of the nucleic acid molecule to a PSNA in the  
sample, wherein the detection of the hybridization indicates the presence of a PSNA in  
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60%

15 sequence identity to of SEQ ID NO: 115 through 217; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 114.

12. An antibody or fragment thereof that specifically binds to the polypeptide

20 according to claim 11.

13. A method for determining the presence of a prostate specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under

25 conditions in which the antibody will selectively bind to the prostate specific protein; and

(b) detecting binding of the antibody to a prostate specific protein in the sample, wherein the detection of binding indicates the presence of a prostate specific protein in the sample.

30 14. A method for diagnosing and monitoring the presence and metastases of prostate cancer in a patient, comprising the steps of:

-143-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the prostate specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of prostate cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with prostate cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the prostate cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

## SEQUENCE LISTING

<110> Salceda, Susana  
Macina, Roberto  
Recipon, Herve  
Sun, Yongming  
Liu, Chenghua  
diaDexus, Inc.

<120> Compositions and Methods Relating to Prostate Specific Genes and Proteins

<130> DEX-0284

<150> 60/252,189

<151> 2000-11-21

<160> 217

<170> PatentIn version 3.1

<210> 1

<211> 227

<212> DNA

<213> Homo sapien

<400> 1

tgtcccatcc attgtagct aaaggttatt ctcaggggct tcctagcact tcaggcctat	60
tctaacatag accaacaaga agcccttcca cagagagctg cagggggttg gtgtgagaag	120
ccaagaacac gtgctgttat ggggagtgc gagtgagggt agggaattga ggaccattat	180
agccactctg tgggagatgc ccttgcgta tattgtccag ccattgt	227

<210> 2

<211> 762

<212> DNA

<213> Homo sapien

<400> 2

accatgtaat gtaaaagtaa gtcgtagaac aaatgtgggc aaactacagc cctttggtca	60
aatccagccc acagacagcc tgttttagta aaatgaattt tatgggagca tagccaaacc	120
ctttgggtta cgtattatct cttcttgctt ttaagctaca gcagagtaga gtagttccga	180
tcagagacca tatgaccagc aaaactaaac atatttgtca tctcagtatt cccattttga	240
gtagatagat gaatcaactc attattttga aaatcaataa taaaaggaaa gaataaagtt	300
ttttattatc ttgttttata ctatagttgg tggacctcag gggtactgag tatttgctat	360
ggagacgttt ttctttataa aaatattttg gctaactgaa tatagaagta ataaaaataa	420
aagaaataat ggatttagaa acttactatt ggtggttgct actatctata tcacaaaaag	480
agaatcacia gcagatctac atgtcttacg gaaggcagtg aatatcacct atgatctcat	540

ttgttaacaa acaaaccaaa aaataacctg gatatgatca agcctccaga tctactactt 600  
atgtcctgga acttaggaag cagacagatt tgaaatgacc attggagaga cagtcctcat 660  
ccaggctcog aatgtgttct gccataaatt ccagcgataa gacttttgtg accccagtta 720  
agatttcggc cttcccgggtt ttccttggtc cgtaaagtcc gg 762

<210> 3  
<211> 665  
<212> DNA  
<213> Homo sapien

<400> 3  
gcggtcgagc tgcacagttg tacggcgagc tgtgctggca taggttgggc aggttcagaa 60  
gtgtaattca ttaaagaagt attttattgg tgacttcttt accaaaaatt tgttaatgaa 120  
attattttca aagtaaattg tgcaaccatg atgtgattag atttaatat taaagtgcaa 180  
attaattaaa atgtaaaatt ttagagctaa tttcataagt ttaacagtag gtaatattta 240  
aattatttta atcattttac ttgggatcat ttcaaatcat acaaataatt acagagaata 300  
ttacggtgaa cctccatgta tcaactccct aattcaacat aatcacgccc tgggaaatag 360  
ttgttcctct atgctctcaa ctaccagta ttgcattatt ttgaagcagg tcagagtcatt 420  
aacattcaat ctgtaaatat tttgatcatg tatatcttaa ggactcttaa aacaattaag 480  
aatatcatga ttactgctgc caaaagtaac aaattatttg atatcaatat atatccagtt 540  
ggatttaaac atagttcata ttaaacatta tccagttcac attgaacagt tgatttacat 600  
caggaatcaa aataaggtct atacattgcc attggttgat aggtctggta agtctcttgg 660  
agtcg 665

<210> 4  
<211> 454  
<212> DNA  
<213> Homo sapien

<400> 4  
tcattggcaa acagcaacaa ttgtatttat acttctctgtt ttccagtttg atgccctttc 60  
ttctcttctc ttgtctgatt gctgcgattt gccattttta caattagggtt atttgatttc 120  
ttattgctgc attgtagttc tatatgaatt ctggatatta agaacttatg taataactga 180  
tatatgattt gcaaatactt tctccatta catgggattt tttgttactg ttgttgattt 240  
tttttgttgt tgaacagaag ttttaacagt ttgatatagg tgattatttt taaaatgtgc 300  
tctgtaaatt ctatgtataa cagtctatct ttggaacttc atgtatgtct caggagaaac 360

3

ttctaactta ataccctgta atatatgtat agttgggtcc ttactgacag gcatttaaga 420  
 ttttctaata ttttgctatc accaacattt cagt 454

<210> 5  
 <211> 573  
 <212> DNA  
 <213> Homo sapien

<400> 5  
 actgaaatgt tggatgtagc aaaatattag aaaatcttaa atgcctgtca gtaaggaacc 60  
 aactatacat atattacagg gtattaagtt agaagtttct cctgagacat acatgaagtt 120  
 ccaaagatag actgttatac atagaattta cagagcacat tttaaaaata atcacctata 180  
 tcaaactgtt aaaacttctg ttcaacaaca aaaaaaatca acaacagtaa caaaaaatcc 240  
 catgtaatgg gagaaagtat ttgcaaatca tatatcagtt attacataag ttcttaatat 300  
 ccagaattca tatagaacta caatgcagca ataagaaatc aaataaccta attgtaaaat 360  
 gggcaaatcg cagcaatcag acaagagaaa gaaataaagg gcatccaaac tggaaaagaa 420  
 gaagtcaa at tacccttggt tgcagatgac atgatcttat atttggaata acctagactc 480  
 catcaaaaaa cgattagaac tcataaacia attcagtata gttgcaagat acaaaatcaa 540  
 catacaaaaa tcagtagtgt ttctatatgc cag 573

<210> 6  
 <211> 632  
 <212> DNA  
 <213> Homo sapien

<400> 6  
 tccatggcga gctcgctcac tttacggcg cagtgtgctg gaatcggctt atatatacca 60  
 ttaagtgaac ttaatgaatc catggacctg tttcagctct tccttacta tagggccagt 120  
 gttttgggta gctgctatga ttgttttgga ctacactggg tggatgattg tattgcatgg 180  
 gattaccata aggaccctgg ataagtgggt atagctgagg gttctgaaat gtaggagaag 240  
 caaaccaca acccagaact atgaattaat cacctgtgag gacaaactgc tttacttcca 300  
 agaagggagg ggactagata aaattaacct tgccaccacc accaagtttc tgtctcagaa 360  
 agttcttgat tatcccttaa tctgaaaaac ttaagaaaat tttggcaaga aggtaattga 420  
 aaattcccat ttggggaaag gggtttcctt tcctgcattt tgagatatta ttctattgtc 480  
 ttttggtttt cattgttgta cattgagaag tcagttgtca gtcttttgca cccctcgaag 540  
 gtaatgagtt tttattatct gctactttta catttaactt ttgcctttgg tttcataaaa 600  
 tttcaatcct ggaatggtgg ttttagaaag tg 632

<210> 7  
 <211> 549  
 <212> DNA  
 <213> Homo sapien

<400> 7  
 acctggcaag tggtaggcat ttgataaata tttcttgagt taatgaatgg ataatgaat 60  
 gcagtgtttt atcaaatagt gggaattaac tggctagctt ccattcatgt atccattcat 120  
 caacaacgtt attgattcca aacaacgtta ttgatcacca atcaacatta ttgattcctg 180  
 aatgcctatt ctgcacccag tatggaccca tattctgtct tctgttagac cccttggcaa 240  
 ttaaaatctg ggattacatt taatgtagta atctacaaag gtaaaactac tggtagtag 300  
 aggacctaga ttaagagatc aactaggctg gtgctctcca atagagtttt ctgtgatgac 360  
 agaactagtc tgtatctgtg ctgtctagta caagcttttt tttttttttt tttttttttt 420  
 taaccggggg cggggaaaac gttaaacaag aatcggaata ttccacaaaa gggggggggg 480  
 acagaaatgt attagagggg caatttcgct aaagggaacg atttaaaacc gaggggcggg 540  
 ggtaaaagg 549

<210> 8  
 <211> 612  
 <212> DNA  
 <213> Homo sapien

<400> 8  
 ggggatagca atgtaattta gctaacttaa ggaaagtgtt tggttaaagt taatgtcaag 60  
 aataagtgtt aaattaaggg atgagtaagt caagagagtt gttgataagt tgagatgaga 120  
 aacgagtcga atgaagtagc agaaaagggtg acatgcttag gttatccaga agtcagaact 180  
 ggaagttgac ggatcaatgt ttaaagaata aggacaatcg atgccatgaa caggatgatt 240  
 aaaaggttta caagggtgta taatgaagta aattactttc ttactttaga tttcagggtat 300  
 aactaggacg tgctttctac gtaaacaaaa agttgtatta ttataatatt tcataatacc 360  
 tcgtatcaca aaccgatgac cagggtgctag ttataaatac tgctagaaat tactgaaaat 420  
 tatgttcatt tttccatttt tgttctttta tttttgaac tctcatttac taacttcaaa 480  
 atattcagtc ggaaataggt aaagaatatg tgcctagagg gaaaaataac agtgtatgtc 540  
 atactgttgt gcaatgagga atccatttcc atcttttagtc tcaaaaagac accttttctc 600  
 gaaatggcct ta 612

<210> 9



<211> 344  
 <212> DNA  
 <213> Homo sapien

<400> 9  
 ggcttagcgt ggtcgcggcg aggtacattt cttaacacat tcgtttggat gggctttgat 60  
 ataagcattc tactaggctt gttaaatgga cagtgttgct tctaaacttt tctcttgctt 120  
 tagattctag aatgccaaat tacctacagt ctttcttcaa tgtttagtag aaaatttcca 180  
 taatattggt caactaggtc tgaaaatttg gaggcataag atctttttct ctctctttta 240  
 ttttaggttc agggaataca tgtgcaagtt tgttacgtgg gtaaattatg tatcgtgcgg 300  
 ctttggtata ctgactattt cgtcaccag gtaatgagca tagt 344

<210> 10  
 <211> 528  
 <212> DNA  
 <213> Homo sapien

<400> 10  
 ggcttagcgt ggtcgcggcg aggtacattt cttaacacat tcgtttggat gggctttgat 60  
 ataagcattc tactaggctt gttaaatgga cagtgttgct tctaaacttt tctcttgctt 120  
 tagattctag aatgccaaat tacctacagt ctttcttcaa tgtttagtag aaaatttcca 180  
 taatattggt caactaggtc tgaaaatttg gaggcataag atctttttct ctctctttta 240  
 ttttaggttc agggaataca tgtgcaagtt tgttacgtgg gtaaattatg tatcgtgcgg 300  
 ctttggtata ctgactattt cgtcaccag ctacaggagc agctgggtctt actgcccagg 360  
 ttctcatgg tcacctcta tttctaata acctcacaat tcctggcaga acagaagtgc 420  
 ctctggtatg accacgtgaa gctcaacaa gagaccacc agagtggcaa cggatctgaa 480  
 atttccatca tatgcagaac agctgaagca actggagagg gtttttag 528

<210> 11  
 <211> 449  
 <212> DNA  
 <213> Homo sapien

<400> 11  
 actggtcagt tgtgtttatt gccttatgtc aacatgcttg tcattttgag aaagtgtcac 60  
 aatcctagaa gaagatgatt tatactcctt gagttcagat tctaactctga gaatgtgcct 120  
 gtgtatcccg tggcatatct ctgtaaaaga tccatggctg cacattgtaa gcatcttgct 180  
 tcttgtagcc tagtccatgt taggaaggct taggttcttt atcttcaaca gcattaacta 240  
 caaattgggt agctaatatg ggaaaatgga agggctctta tcacctgtag ttttactttt 300

6

tcttttccat agtcttcata ttgttgacat atttagggta atccagttgc tgaaggacat 360  
ggacagaaca cagaattggt atcaggattt gccacaggt aactatttga tgctgagttt 420  
aaactccttg agtcttagtg taagccgaa 449

<210> 12  
<211> 1036  
<212> DNA  
<213> Homo sapien

<400> 12  
actatgcatt ttatacaacg tgctattgtg aaaaatgtta catctttcaa tataacatat 60  
atactatata atgtaacaag atagacataa tataaccaaa aggaaaaatt aaccattagg 120  
cttatggctt aggacgagag acacatggta acgtgacagg caccagggat acaagaacca 180  
gtgtagaact gtgttaataa agccttgccc cgtcaatcta caaaactaac gacaaagaca 240  
atacaggctg gctttttggg tgcatggggc ctctgtaaaa ctacttcaaa gtcagctgtg 300  
catatcaagt cccttaagaa taccatggg gactgaacag gttctcgtgg tccaagcata 360  
ttctggcgaa tccttcagat tttatttccc acagaacttt ggaatgtcaa agaaaccagc 420  
atatgaacag gactctgaaa aatccttaag taaaggaata cctttgcatg tgcttatatc 480  
cacattttcc caatgttggg caacagaacc cctggaatag taactatcaa acttctggaa 540  
aattgccaca acttccatga atgcacattg gggatttctt atgcataaag gccttgagaa 600  
atattctgct gaaaaactat actctcgcat actgggtata attttgatga tcgttcaaca 660  
aaattatcac tcattccaaa cctgagagga agatattctt ttgccttttt atcagtgaga 720  
caataatctc aagaaaatta agtgacaaa cgatgattac cacagcctgt atacctgctc 780  
attcctccta taactgtaag gggcagccta tcaatcatag tattattata cttcttattt 840  
atgaaatatg aaggatgtat ttctttttta accaactc tgttttatag aagtgttgac 900  
attttgaagc tcgcttagtt cactcccagg cttgctgcat cagtgccagt ttctccaact 960  
tagtctgacg acacatgtaa gccgatgtcc agcacactgc ggccgtataa ctggcaggct 1020  
ctgccgttcg tcccc 1036

<210> 13  
<211> 988  
<212> DNA  
<213> Homo sapien

<400> 13  
acttctgctt tatctttgag gagatgctgc caaatttcct cctgtaataa ccgtaaagg 60  
ttgttttttt ttttgagggt tgggtggcttg gattccactc tcattctctc accaattatt 120

```

gtgtggtgtg ccaaagcacc atactaatgt gcggtgtgaaa agagagatga ttagaagaga 180
cacagtctcc gggttggtcac agactetaag acgccttccc gtgctcatat gaggaagacc 240
atagaatata ttccaggggtg agggacagtg acaccatatt ttatgaaagc acttcaatgt 300
gcttccccat tcccagccta aagataccca ctttagaatac gtgcggaatg ggaatgggga 360
tagccatata ggggttcctt ttaacctaaa aatttacggc cctaggttgg gaaatttttt 420
tcctccaggc aaccttgtcc ctaattttat attaaaggtc cttttagaac caaatttaac 480
agaacacata acctgtgaat tttaggtttg ttagagaaat aggttgaggg ttttggggga 540
ttctacaggg agcccattac atatcaaaat tgtatcattc atgtatacat ttcaagaggc 600
atgtgcctat atacactact aataactcat ctagaatata agattccttt tgaactggca 660
gactccaata tggggaataa agtgatgctc gtttgccaa tctttataaa acttggttag 720
gttttatgcc ttccaaact atactggacc agatgtctcc ttcctcttgc tttctgccc 780
tcagggtctg ctgtagtgta gaaagcactg ttcctctgaa tttcctattt acctgttttg 840
gcctgtagtg ataattattt ttctctgaag tcgcacgatg atgtaaccaa ggagggacaa 900
tttgaggctc agtagtatcc acatgacaca gtaattacag gcaactgctg attgcttttg 960
acctgcccaa cgccatcgca gcatcatt 988

```

```

<210> 14
<211> 499
<212> DNA
<213> Homo sapien

```

```

<400> 14
gacacaatgg tagccagtaa gcgacctttt ccagacatt gaataacact ttaatcttag 60
acttcccagc ttcaagaact atggtaaata aatttctatt atgtatagct ccctagtttg 120
tattgattta ttatagccta aatggactaa ggtagtcttt gtgttctacc tttctgcaa 180
gactatacta ttttggttat tatagcttgg aaataagttt tgacagcagg gagtctgagg 240
ccctccaaat acattctttt tcaagatcat tttggccatt tgggggtcct tggaaatgtt 300
tgaattatta gtatggattt ttctatttct gcaaaacatg tcattgaaat tttgaaaagg 360
attacattaa atctgtagat tgctttgggt agtattaaca tottaccagt attgagtctt 420
ccaattcata atcataggaa tctctcccat taattatatg tottaattta tttcagcaaa 480
attttgcagc tttcattgt 499

```

```

<210> 15
<211> 888

```

<212> DNA  
<213> Homo sapien

<400> 15  
actgtgtaag aattggccgt cgattaatat tattcattgg ataaatatca aattcaaaat 60  
accgtttaca ttgggaaagg gaaaaaggag agaaatatat gagagacgta tgctgggggt 120  
atctacgatg tttttttttt ttgatttttt tatgtcattt taaatgtgtc tattcttctg 180  
ttctgttctt aaagtatacc tgtaattatt aggattttta ttatggattc tgtcttattt 240  
attttttaat cttctttatt gtgtctatct atttgccttt ctccatattc tgggagaaac 300  
atctaagttt ccattcctta taactgatta tttttcttgg gatcaattct ggtcttcacc 360  
atctgccaat gtgagtttta gtttctctat aaacttttaa actttaaact tgttttcttt 420  
ttattatatg ggggggggca atgctttctg taacccctt ccagtttttc tcaggctatt 480  
cttgtttggt ctttgttttc ttttctgtt tcattgagat gcttacaaaa tggaagatgg 540  
ccaacatttc tggaaaaatt ttgtttcctg aacatagaga aacgtgttag aaggtcagg 600  
tgctctgaga ctgtgttggt aatttctttt tttttttttt acttgtttgt cctttctttt 660  
attatctctt gctgttatcc tcactttttt ctttgcgtgt cttcatcctt taattagcat 720  
tcaaactaat cgttattaga gtgtgtaaca tgatctttga catacctagg taattttact 780  
taagtatatg agttaagttg atacgcacag ttcctcagca tgaactgctc attgttttgc 840  
aaatttttagc tgaagggtgac atcactgtac ctggccgga ccacgtaa 888

<210> 16  
<211> 669  
<212> DNA  
<213> Homo sapien

<400> 16  
ccgtgacgag ctctcatcac tattaacgag cgcagtgtgc tagaactcgg cttacatctt 60  
atagctttcc cctgtaattc cctccactta tgttgtcttt tgtgattagt tgttttatga 120  
gtagtgacaa atgttgcttc ctttctcatt cttctttggc tatacatatt ttcttgggt 180  
gttaccatgg tgattacatt ataatatccc aaagttataa caatctaatt tgaattttga 240  
atatcaattt catacaaaac tctacatggt ataccacttc tgtttcccat cccacttta 300  
cattattgggt ttgaataatt atatctatat gctgttatat tcaactaacag atttataatt 360  
acttcttatt catttctttg aaatcctata gaaaataaaa agtcagcta tgaacaaaa 420  
taatactgggt ttttaattgtc tgcccgctca tttaccttta ttaggagagt ctttatatgt 480  
tcagggtggct tggagttaac tatcatcttt ttattataac tcaaggaggac tccatttatt 540

atgtcttgta atcaggaata gtggcaacaa gtcacctggg aatgtcataa cttccccctc 600  
 atttttgaag gacagttttg ccagataatt cttgggttact agttttttct ttctgtacct 660  
 cggccgcgc 669

<210> 17  
 <211> 566  
 <212> DNA  
 <213> Homo sapien

<400> 17  
 actttggatt cacttctggg atacatagca gtgtctataa atacctctca tgtccagaat 60  
 agaaagcctg agggagagat agaggctaga tatttgatat taaaaaaaaat attccaattc 120  
 taccttttgc acagaattga atatttgtaa ttgtatcttg tagatatcaa attaaaagca 180  
 taagtttcat tttaacattt ataatagtat atcatctatg gagaacagac aatacatatt 240  
 tatattacac acattttatat gttcctaata aggtgtcttt atttagtaga caaatgttga 300  
 acttttcgca taaattaaaa attattactc aaggaccttt attgggaaat tccatatttg 360  
 tottaaaaaag aacaattatt ttataacgta attcatcact cccagattta aaagctttca 420  
 acatttaggg gacatgggat attgtattgc actttcttaa aataaaaaaac agtatcttca 480  
 ttttttacac aggaatgttt gttagcacag ttgtgatggc tcatgcctat aatcccaggg 540  
 ggtggcagtc caagtaactc gggcgg 566

<210> 18  
 <211> 721  
 <212> DNA  
 <213> Homo sapien

<400> 18  
 cgcccgcgta gggtattatc tgtctctgcc atatatacgg cttacaaaga agagctgggtg 60  
 ccaatcttgt tgaaactatt caaaaaaaaa aaaaaaaaaa ttgggggaag aggtttttct 120  
 tacctaactc ttttttatga aaccattttt ctccccttga attaccaaaa tctcaggcaa 180  
 ggaacacctt gcaaaaaaaaaa aacttacagg ttataatggt ctctaattga ccaattagaa 240  
 ttgttaaaat tctccttctt agaaaaaatt atttagccaa gacccaatat ctccgaatta 300  
 accatcatte aaaacagggt ttagttttat taattcatta cagggtgtggg atttattatt 360  
 ctccaggggg acgccagggg tattaggttc caggcatctc cacaaaattc acttaacgtg 420  
 tgggtgattac attcaccatt ataacaggaa cttataaaga caaaaacca tgtgagtcac 480  
 tcttcaatta gatgtgcaga agaagagcat ttcaattaa gggccataac atttcttttg 540  
 attaattaa attcttctca tacaattct cagccaaatt tagggcttta atggtaagac 600

accttcacaa aaattaagag ccattttatta acaaacctta gagccagga taactactag 660  
 gggggcgaca cgctgaaaaa tttccctata caagcgggtgt tagacaaaga tgacccatct 720  
 c 721

<210> 19  
 <211> 1053  
 <212> DNA  
 <213> Homo sapien

<220>  
 <221> misc\_feature  
 <222> (46)..(46)  
 <223> a, c, g or t

<400> 19  
 atagatgctg gatattagac ctttatcaga tgcatagttt gtaaanattt tcttcccttc 60  
 tgtagtgtgt tcaactttgt gatagtttct tttgctgtgc agaatacaga atgataaggg 120  
 ggatattgcc actgacccca cagaagtaca aacaataatc agagaatatt ataaatacct 180  
 ctatgcaagt aaattggaaa atctaggaga aatggataaa ttcattgtaca tacacccttc 240  
 caagactgaa gcaggaagaa attgaatccc tgaagagacc aatatcatgc tctgaaattg 300  
 aatcagtaat aaatagccta ccaaccacaa aaagcccagg accagacgga ttcacagctg 360  
 aattctacca ggtatacaaa gaagagctgg taccattcct gttgaaacta ttccaaaaaa 420  
 aaaaaaaaaa aattggggga agaggctttt cttacctaac tcttttttat gaaacccatt 480  
 ttctcccttc gaattaccaa aatctcaggc aaggaacacc ttgcaaaaaa aaaacttaca 540  
 gggtataatg ttctctaatt gaccaattag aattgttaaa attctccttc ttagaaaaaa 600  
 ttatttagcc aagacccaat atctccgaat taaccatcat tcaaacacagg gtttagtttt 660  
 attaattcat tacagggtgt gtatttatta ttctccaggg ggacgccagg ggtattaggt 720  
 tccaggcatc tccacaaaat tcaacttaacg tgtggtgatt acattcacca ttataacagg 780  
 aacttataaa gacaaaaacc catgtgagtc attcctcaat tagatgtgca gaagaagagc 840  
 atttcaatta aagggccata acatttcttt tgattaatta aaattcttct catacaaatt 900  
 ctacgcaaaa tttagggtct taatggtaag acaccttcac aaaaattaag agccatttat 960  
 taacaaacct tagagccagg gataatcact aggggggcga cacgctgaaa aatttcccta 1020  
 tacaagcggg gtttagacaa gatgacccat ctc 1053

<210> 20  
 <211> 631

11

<212> DNA  
<213> Homo sapien

<400> 20  
ttgacaagag attaaaaaca accaatgcc a tattaccaag tgaattat ttt actttgtagg 60  
agaaatgagg taaagggata cattacttct ggccaatagg aagttagact aattaccaat 120  
ggcaattaac atcttgtctt ctttgaaga caagacatgg tgcaaaagga aataaataat 180  
tctaagtctt gttaagacta tcctaaaggc cattgccaga agctattaaa aacataatct 240  
taaaaatatt cagtctattg gaatatatcc ctaattctaa ttaaaagctg aataaacttc 300  
ttgctagtat taattagttg tgggtgtgaa tttcatcagg tccccatcgt attaatagaag 360  
gacattcata aaatagatag taaagataaa tttaaaagcc aaccagtggc ctcaccctct 420  
tttttactca aaatatagtt tatattaata aactggaaag acatcagaag tgtagacagc 480  
tactgtaaag taaatattaa atgtaggact caatcaacaa gaataactat aaagaatatg 540  
atgccaaaat aatatgacac gaataaccct tcaatgatcc attaaaaatg tggctggttg 600  
acaaagaaca aattaatggt ttattattag t 631

<210> 21  
<211> 888  
<212> DNA  
<213> Homo sapien

<400> 21  
ttgacaagag attaaaaaca accaatgcc a tattaccaag tgaattat ttt actttgtagg 60  
agaaatgagg taaagggata cattacttct ggccaatagg aagttagact aattaccaat 120  
ggcaattaac atcttgtctt ctttgaaga caagacatgg tgcaaaagga aataaataat 180  
tctaagtctt gttaagacta tcctaaaggc cattgccaga agctattaaa aacataatct 240  
taaaaatatt cagtctattg gaatatatcc ctaattctaa ttaaaagctg aataaacttc 300  
ttgctagtat taattagttg tgggtgtgaa tttcatcagg tccccatcgt attaatagaag 360  
gacattcata aaatagatag taaagataaa tttaaaagcc aaccagtggc ctcaccctct 420  
tttttactca aaatatagtt tatattaata aactggaaag acatcagaag tgtagacagc 480  
tactgtaaag taaatattaa atgtaggact caatcaacaa gaataactat aaagaatatg 540  
atgccaaaat aatatgacac gacaccactc ctatagatca gtgcgggggca atgcactaga 600  
aaagaagtct aactatgaag tattagaaaa agatgttggt ttaaagcgat ttttgcttaa 660  
gagtttactg gattctgtca gggccaaaac actaagaaaa ctgatgcaac aaacatgtag 720  
acaagttacg aaccttaata gagaagaaag tattctgaaa ttctttgaag atcctgtctc 780

12

cagtctacag atttggataa ggaatgcttc aagtgtgctc ttggttcaag ctggattatt 840  
 tcagtggaac tggcaatcgg cccagaagaa ggaatcagtt acttgacg 888

<210> 22  
 <211> 363  
 <212> DNA  
 <213> Homo sapien

<400> 22  
 accgtgcaga tctcatgcta tgttggttac tattttctat ctaattctga aaagttctgg 60  
 aataataatg tccatttatt tgattttagg gatgtttcag attcatttcc aggagtgggt 120  
 cagccatagt ttattcacat actgtattca aataatatta gacttaataa tttcaaaaat 180  
 acatatttag gcctctgctg tatgagtaac agagataatc ttttaatttcc cttcccttcc 240  
 ccacagagca cttggtgtaa atggaatatt tggctctgat atgtctcact ttcagtagtt 300  
 tgaaagtatt taacaaagaa gataaccatt tggatatgaa gcaaactatg cctggtgctc 360  
 ata 363

<210> 23  
 <211> 383  
 <212> DNA  
 <213> Homo sapien

<400> 23  
 tgtgctgcag ttcggttacg tggtcgcggc gaggtgtcaa gctgattgat aaggttgctc 60  
 aagtgggtcta catagctgct gattttctgt cacaattatt gattgagttt gttgaaattt 120  
 ctcattataa ttgtggattt gactattcct cttttagtct cttccagttt ttgatttgtg 180  
 tatcttacag ctctgtcatt atgtgcatat gtatttcgaa ttgttatgtc ttcttgatag 240  
 ttaacctctt taatcactgt aaaatgacct tttttatcct cagtaatatg aattgttcca 300  
 aaatatactt tttctgatta tttaaatagc aacctccaga ttatcttata tggttggttt 360  
 tatactttcc tacatccctt caa 383

<210> 24  
 <211> 711  
 <212> DNA  
 <213> Homo sapien

<400> 24  
 acactgagac tgaacaggtg ttaataataa aatgccagaa caagagacat ttactgggac 60  
 tgtcccagggc aagcaaagat agaagatcac catcactaag actggcctag gttttatcct 120  
 caattatggt aactctataa tgtcagtctg gcacaataac ttttaggctt aattttctca 180



13

tctaaacata gtatgggtat aacaacgatg atcttaccce tattacaaaa tttttttgag 240  
 aatcaattga tgaaagcatt aatctggtga agtttttatt attcaatgat ttagggaatg 300  
 ttttcacata atggaataat tagtaacaac acaagcagtc atggcataaa agagtaatga 360  
 gaactaacat tgtcttaact cgttacatgg tggtgaggtc cgttatattt aataactaacg 420  
 ttttacattg ttatagtatc tatttgtaaa tttatgactc attctttcca tactagtgggt 480  
 gaggttggtt tatgctaaac ttaaaagatg cagagaccta agatcccagg gtagtaactt 540  
 ggctacgatac aactaacta tgatagtgtt tataaataac ccagattcta caccaagaag 600  
 tctgactcaa gagctcataa gtgtcagtag gggggtgtag tcgtattata tgctgtctgt 660  
 actccttctc ctagtttggt taaaaaacac ttacatgtac ctgccaagc c 711

<210> 25  
 <211> 812  
 <212> DNA  
 <213> Homo sapien

<400> 25  
 ctgcacggct cgcctcacgt gtaacggcgc agtgtgctgg acttcggctt acatatgaaa 60  
 acattttcat accttctggc caaataattt aatggtgccca ttttttataa agatgaaaaa 120  
 gcgagagaaa ttaataaaaa gaatgggatg ttgctcaaac cattatgttg aggttaataa 180  
 aactaatttt aaataatatt ttaagatgca cagaaataaa attagaggtc gcattcaagt 240  
 tattcattta taattgtaaa taatttaact acaataattg tttagttacc ttatttcaat 300  
 ggtatgtccc aataattaga cataatgttc caacagaaat taaccaaga aggcaagaag 360  
 tctcaaaagc atattataaa taatacagtg tgtaatttaa taatacataa tgaaaacatt 420  
 aaccatctaa ataatgaaac attactttgt aaccaatta tattaattaa taaaattctc 480  
 taagttcaat agtttattat cattaattgt aacacagctg gttgaatgaa attttgataa 540  
 tttccattag tgcactttta actttataag aacatataga ctagacttca gtaagatgga 600  
 agggtaggat ttttcacctc tcattctagc ctacaaaaca acttacttgc agctattcag 660  
 aagtggtaat accccttggga gaactctaga tccactccag catctgtagt gtagaacaac 720  
 agcagcagat aaaaaaaaaa aaaaaaaccc acgttttagg ggggaacacc ttttcctttt 780  
 ctgggcggtt tttcccgggg gagagtcctt ta 812

<210> 26  
 <211> 440  
 <212> DNA  
 <213> Homo sapien

<400> 26  
ctgtccatgg tagaaatggc aagggttttg gaaacacaga cttggactct aatctcagtc 60  
ctttggcata ctctgtcatc aaggcaattt acctaaacat tctgaaatgt aggtttcaca 120  
tctgtaaaat ggggagctgc tgcagctcac agtatgttgt aaaattaaat gagtatataa 180  
ggcacggtac atgtaactgt ggaaatgctg agttgcaggg tatgcatatc ctcaagttta 240  
atggatatca ccaaattgct ttccacatca tcaaaattct taattataag caagagaaca 300  
ccattatgga tcattcaaac caggaaaatt tcttttttta aagaatgttg gaattagctt 360  
acagaatctc aggaagtgc tgcagaaccac ccaaatatga atggaattga ctaatggtga 420  
ttgctgcact ggttatgggt 440

<210> 27  
<211> 164  
<212> DNA  
<213> Homo sapien

<400> 27  
ccctcactcg acactgtaca cactatgcgc ctggggctta tagatgcatg gtcgagcggc 60  
ccgccagtgc ttaatgtagt atctgcagaa tacggcttac aataaaccgc catgacacat 120  
gcacccctga acctaaaata aaagttaaat aaagtcacat ttga 164

<210> 28  
<211> 186  
<212> DNA  
<213> Homo sapien

<400> 28  
acagttctgc ctattcttaa ataactgcct aattgtctat ttaaaattac attattgaaa 60  
gggtgtcttca agcaactctc aggtgaaata gcctagcttc tgttgcgtga gaactgcttt 120  
cctggagcct tcattcagtt tgtagctttg tagttgatac ttcaaattatt ataagatcac 180  
tggtgt 186

<210> 29  
<211> 186  
<212> DNA  
<213> Homo sapien

<400> 29  
actccagcct ggggtgacaga gcgggaccct ctaaaaacaa atgaataaat aaactccggt 60  
tccggaaaaa gaatgctggt aagagaatga aaagacaagc gacagactag gagaaaatat 120  
ttataaaaaac acctaactgt ataagggact ggtatccaaa acatacaaag tgttcttaca 180  
gctcag 186

<210> 30  
<211> 692  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (228)..(305)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (580)..(580)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (587)..(587)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (655)..(655)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (658)..(658)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (677)..(677)  
<223> a, c, g or t

<400> 30  
ctgagctgta agaacacttt gtatgttttg gataccagtc ccttatacag ttaggtgttt 60  
ttataaatat tttctcctag tctgtggctt gtcttttcat tctcttaaca gtgcatttca 120  
catagcagtt ttttaatttta atgaagtcta ccttatcaat ttttctttca tgaattatgc 180  
ttttggtggt ctatctgaaa actcattgca aaacacagtc acccaaannn nnnnnnnnnn 240  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
nnnnncacat tctttgaatg ccaaactg acaaggtaaa actaaagaga caatgagagg 360  
atccctatta agtggctctc ttatagaagt cacagtggct acgtgtggta tttccccaat 420  
cagactcatc taacctttca aaattaaaca accaaagaaa tcaggaaatt ccaaagagtg 480

attttttcttt ccaaggggga gaatccaaga agtgatactt ttacattcc tccaatttgt 540  
 ttggcctccc tatttatgga aaatatattt atagcagtn cattgcnaaa ttccctcaac 600  
 atcagtagga ggaaggcctt atcaaaagag cacaggctgg gcgatgtcaa aaggnggnag 660  
 tgatgatgaa tagtacnaaa tgctccttaa ac 692

<210> 31  
 <211> 530  
 <212> DNA  
 <213> Homo sapien

<400> 31  
 cttaatttgt tcctggagcc atttaggatt atttttgggtg gacgctatca aactatacta 60  
 tgaagagaga gaggtttggg ggaacactat tcccaccttt tccacacaaa cacaaattca 120  
 accctatttt ttagtgtaat atttgtatc ttccctaagtt ttttttgggtt tgggttggtt 180  
 gttgggttag tttgggttct gataaagtgt ttcaaggcca aaatgtttga agactactgt 240  
 tttaatgttt tattaactag ttataactag ttttaataact ggtttattag cagcctgggt 300  
 tatttgcaga attcctgttt aagatagtat atacctctaa tccaaaagct atccattgta 360  
 tggaatatat agccttccca gttttttttt ttttagcttc attcatgcat gcctgcctgc 420  
 atgtatacta gactcaggac ccctaataca aaaactatcc attgtataga atgtgtagtc 480  
 ttccaatttt tttgtacctc ggccatactg catctgcagc acactgcgcc 530

<210> 32  
 <211> 663  
 <212> DNA  
 <213> Homo sapien

<400> 32  
 actaaatggt aatcggcaac catcaagcat aactgggttaa gaatggaaga agggtaaagg 60  
 ttgtgtgggg ctacattgat ttcaatcaag gtaatgaaac aagcaaaaaa aaaaaaaaaa 120  
 aggaaagaaa gaaagaaaaa gaaagagaga gagagagga gggaggaggg agggaggaaa 180  
 aaagagagag gggggccggt gacagaaggc aaattctatt ggtgggggag acgggcaagt 240  
 taaaacaaac acggggggtt ttagttatat aaccctgaa tactatggga gcatatacta 300  
 ggggttctta acatacatga agatcagc atctatacag attttcactg agtgacatgc 360  
 ccttgaactt aatgggtcac aactagtcct tcatttactc ccaccaaata tccatttata 420  
 gtgccactcc taatgtacat aattaatgtg aaagtgacc tactttataa gttttttatg 480  
 cagacaaaca aaattcagct actgaattac tctgatgcca tcaacaaatg gtgataacag 540

17

ccaagtaggc agatgtttta ttaccctaaa catgttgtaa acaggcttca cttggcttgg 600  
tgtttctatg tttgattttg aggtatttta agaatgtcac tagccatgct acgtccagta 660  
agt 663

<210> 33  
<211> 694  
<212> DNA  
<213> Homo sapien

<400> 33  
acttttaaat actgtaaggt ctgaaggtag acaatgcagt attacacaaa ggaagaatat 60  
ttaatggatc cttttatttc caggggagca gagtgataaa agtctatttc tgtgatactc 120  
attattttgc ttcttatttg tttaacactt tatttttaaaa aaatcaaaag cagtttttga 180  
aagaactaca gacagacttc ttggccccta aatacatcaa tgaatcaatg cctagaactg 240  
gatttcttat aaaatataaa ttggcgttta ttgccacaga ggctaaacta taacaaaaaa 300  
tgaaagaggg tgggaaaaaa aaacagtttc aaagaaagtg aaggcaaata acagacattg 360  
ttgggtaaac acattattgc tactaagctt tctacaatcc accttttaag aaaaatgtta 420  
tattttcatt tgtcacatca aatgcatttt tttttaacaa tgttggtttt catcaaaaga 480  
aacacctcag actgatcaag actcagctcc atactaatcc agttcctggc ttctgcagaa 540  
actgaatata tcttcgatgt gtattccaga aaaaacaggc catttttatcc aagaccaaga 600  
gcaccccaca aaaaaacaaa agcaaagga aattttotttt gtttttgtaa gtcaattcaa 660  
gacgagaaat aatatgcctg catatggctt tagc 694

<210> 34  
<211> 564  
<212> DNA  
<213> Homo sapien

<400> 34  
acaaggagtt aaaattaact tggaatttgt taatttagac acaaaacttg cataatacat 60  
atctagaaat cctttgacct atagtctata cgtagaacat atatgtgtcc ataatttgca 120  
cacatacagt gagtattaaa tagcaaaact agattacttt attctgaaat acaccttcac 180  
tgagagttaa agtattaata attccacatt tatgtttcag aagaaatcaa gaggttcaca 240  
aatatccctt aagaaatatt ttactactta tttctttctc caaatttgct acatggaact 300  
gtgcattatc attcatatga attcacaatt tataacctat ttgctctaaa gaattcatta 360  
caatttacgg tatgaatgga aactaaacat agagaaagtg cctaaacact acacattgat 420  
tcaatggata aattttttat tataaaataa attattcagt tcatggtttc tgacaaaaat 480

cagatcctcg ctatcatata tatattaaat acactattaa aatccaacat gccatgtaat 540  
gtattcattc tggattccaa cagt 564

<210> 35  
<211> 639  
<212> DNA  
<213> Homo sapien

<400> 35  
cgattgtaac cgctcatatg ggcaatgggc cttcaagagc agctcgccgc cccccgtgtg 60  
atgatatcgc cagattcggc ttctgagcgc cccccgcgc agtaactatg gtatctttta 120  
tgatcagttt atagtaaacc cttgcataat ttgttgacaa attttctaata gcttcaaaat 180  
gtttggcttt caaagtttta attaacataa ttaaataaaa tgccattctt tgacaatcct 240  
atgcttagaa atgggtacta tccccacaac tgaactgtct tgggccaatg ccaccatcta 300  
gtggctaata cccagaatca cacctctgtt tctcctttta tacctttgct ggaaattcat 360  
atgcatgggt tctgtttttt agagtgggtga ccaacttaga cttagagaat ttcattgtcac 420  
tggtagtctt agccctttta ttcattctct tctaatttta tcttttattt ctctattttt 480  
atttgcaact caacaagatt tccatgtcag agctaacata ttttgtgaat aatttaataca 540  
attagttttg aagtaaaatt agatgttgta aaatggctct gaacttgta aactccttga 600  
aggccctcag ttatttccta tagtatctcc atttctggt 639

<210> 36  
<211> 871  
<212> DNA  
<213> Homo sapien

<400> 36  
ctttccagaa ggtgtggcaa ccaagtcat gagacaaatc cattagaaat gttgaggttg 60  
gacaatacat tagaggaaat tatatttaag ctggtccctg gactacgaga acaagaactt 120  
gagcgtgaat ctgaattttg gaagaaaaat aagcctcaag aaaatggaca agatgatact 180  
tcaaaagctg acaaaccgaa agtagatgaa gaagggtgatg aaaatgaaga tgataaagat 240  
tatcacagaa gtgaccacaa aattgctatc tgtctagatt gtttacgaaa taatgggcaa 300  
tcaggggaca atgtagtaaa ggtgagtga caagtactat ggtatctttt atgatcagtt 360  
tatagtaaac ccttgcataa tttgttgaca aattttctaa tgcttcaaaa tgtttggtt 420  
tcaaagtttt aattaacata attaaataaa atgccattct ttgacaatcc tatgcttaga 480  
aatggttact attcccacaa ctgaactgtc ttgggccaat gccaccatct agtgggtaat 540

19

gcccagaatc acacctctgt ttctcctttt atacctttgc tggaaattca tatgcatggg 600  
ttctgttttt tagagtgggtg accaacttag acttagagaa tttcatgtca ctggtagttc 660  
tagccctttt attcattctc ttctaatttt atcctttatt tctctatttt tatttgcaac 720  
tcaacaagat ttccatgtca gagctaacat attttgtgaa taatttaatc caattagttt 780  
tgaagtaaaa ttagatgttg taaaatggct ctgaacttgt taaactcctt gaaggccctc 840  
agttatttcc tatagtatct ccatttctgg t 871

<210> 37  
<211> 188  
<212> DNA  
<213> Homo sapien

<400> 37  
tttttttttt ttttttcccc ccggaaaaaa tgttttaatg tccctttact tggaaaaaaa 60  
atccaataat acaacgagtg taaatttttg ttccctcagaa aaatccatct ctataacacc 120  
gggtgggttcc agccgcagtt atattocacc attggccaag gtgagggtca taaaattgtg 180  
gggtgggg 188

<210> 38  
<211> 419  
<212> DNA  
<213> Homo sapien

<400> 38  
ttgatttaag acttctaagc ctccagtatt ttaataaata aagtcctaata taatagaaac 60  
tattcaattt atgatacttt ggactaaagc aggtcagttc tctaacaaga cgtagaactt 120  
cttatcatgt atgcacctaa tactataaga ttgcataaaa atatgttgag caaaaactga 180  
aagaactgca aggacaaata gaagaatcta ttgtaataat tggatacttc atcacccctt 240  
taccaaaaac ggacaaagtg agcagggaga aaatttgtga ggacatagct gaactcaata 300  
gcaccatccc accattggca tctacaggct acttcatcca acagcagagt acctgccca 360  
gccgaattgc agcacactgc gcgcgtatta gtgaatggag ctcgtacacg cttggattc 419

<210> 39  
<211> 358  
<212> DNA  
<213> Homo sapien

<400> 39  
ggtacctctt actcagtgat cttcccctct tgcaggagtg atctcctctc cagtttctgc 60  
ctgcttttgt tatgtccca atgtctttaa gaaggatatca tcattataat tatagcttat 120

20

cttggatatca atggaaggtt aatctgacat aatgttttct ttcattatca gaaccatacc 180  
tattctgaat gaaattaaac cctcaaatg attcaaatct atctctgtag ttcactacct 240  
aaaccactgt gaaactgaca ctgagcacia tacatttagt gtctagggaa cagtttcttg 300  
agattaaact ctctttcttt tcacaacagc atatctaatt caatatcgca acaaatgt 358

<210> 40  
<211> 421  
<212> DNA  
<213> Homo sapien

<400> 40  
actcaaggag gttatttttag tggaagcagg ctggtgggtat gggacagata aagtcactgg 60  
ggtcagatga ccaaatgact aggtccatct gtaagaccat tctaaacttt ggggaatctt 120  
tccccatttt tactgcctgg attccatgag ggtcagttat caggtctttc ttttataaag 180  
gaagttcaat atattcttgt atagggtaat actttgttgt ataggtttct aatagttaat 240  
ctgtgctgag gctggccttg tttcctaata gaatcccagg gccaaaggta aaaatcctca 300  
ggccaagtta catcactactg ctgacatata gtattggccc tgcaagatat ttaactgggt 360  
cttacggcta ccatttatag ccacgctaag tctcttagag tatactacaa gcggacaatg 420  
t 421

<210> 41  
<211> 201  
<212> DNA  
<213> Homo sapien

<400> 41  
aaaaaaaaacg aaataaaaag acaaccaatt ctctcttgt gttcttttcc attgaagaac 60  
gcaaagaaag gcagtgcagg gcagggattt gacgttggct tgatagtgg cagcttattg 120  
ggcttgctcc aattgaccaa gggactcatt aggcaaggag gctcctcttt cgtgcctaca 180  
ttcttccacc ccagggaagg t 201

<210> 42  
<211> 814  
<212> DNA  
<213> Homo sapien

<400> 42  
tgtgctggca attcggttta caaaataatg atattaaaga agctcagtga agccagtgtg 60  
gttgacatt cctgtagtcc caggctatgc cagaggttga gaagcaggga ttacttgacg 120  
cccaagaggt caagtctacc ttgggcaaca tagtgagaca cagtatctaa aataataata 180



21

ataataaaag ttcagtgaga tacgagagaa cacatgataa acaacacaaa gaaatcaata	240
aaagaaattc agaataataa tgtgaaatth accaaagaaa ttggtataat aatatagaac	300
aaaaaagaaa ttctacaact aaagaattca ttgaatgaaa gaaagtaaatt attggtgaaa	360
gttttgacca tagactaggc caagtagaat aaagaatttc aggaatttaa tgtaggctct	420
ttgaaataac ccagtcagat aaacatattt ttaagaattt tttttaaaaa gagaacaaaa	480
cctatgtggc atataggaaa cataaagcaa ccaaatattc aggttttcaa tgtctcagaa	540
ggtgacgagg acaccaaag ttgtagaaaa ccttattaac agacacaata acccgaaaat	600
gtcttggaat tgtagcaaaa agatgcagac ctcccagtta cagggaaact tcagaaataa	660
tagatccaat ttccaaagat cttcttccca ttgattttcc ttacattttc caaggccaat	720
tagagagatc ttaaacacaa aaaaaggctc ttcccttagg gatccctaatt aagggtttct	780
ccaacttcga aacagggttt tctctctgca aaaa	814

<210> 43  
 <211> 1205  
 <212> DNA  
 <213> Homo sapien

<400> 43	
tttttgcaga gagaaaaacc tgtttcgaag ttggagaaac ccttattagg gatccctaaa	60
ggaaagacct tttttgtgt ttaagatctc tctaattggc ctgggaaaat gtaaggaaaa	120
tcaatgggaa gaagatcttt ggaaattgga tctattattt ctgaagtctc cctgtaactg	180
ggaggctctgc atctttttgc tacaattcca agacattttc gggttattgt gtctgttaat	240
aagggtttct accacttttg gtgtcctcgt caccttctga gacattgaaa acctgaatat	300
ttggttgctt tatgtttcct atatgccaca taggttttgt tctcttttta aaaaaaatc	360
ttaaaaaat gtttatctga ctgggttatt tcaaaagacc tacattaaat tcctgaaatt	420
ctttattcta ctggcctag tctatggta aaactttcaa caatatttac tttctttcat	480
tcaatgaatt ctttagttgt agaatttctt tttgtttcta tattattata ccaatttctt	540
tggtaaattt cacatttata ttctgaattt cttttattga tttctttgtg ttgtttatca	600
tgtgttctct cgtatctcac tgaactttta ttattattat ttttttagat actgtgtctc	660
actatgttgc ccaagggtgt cttgaactcc tgggctcaag caatcctccc acctcagcct	720
cccatgtagc tgggactaca ggcattgacc tatgcttggt ctttagtgcc ttaatatata	780
atttacacac agcttctgag acagtgtctc ttaataacac agagtgcgt ggcattgtact	840
ttaacagtaa ctttaggttc acacagtaaa taaaatctca caatgccaaa attttgaaaa	900

22

taggactaaa acaaggtaga tgtgtgtagt ttcaaaagtc tcaccactga actcatccta 960  
 tgcaaattcc taaccaatc atgctgttc caatagggc cagcactt ctcatgtcca 1020  
 ttagatgatg caaaaaagga ccaaatggct aacactccct tattttcaac aagaagtaca 1080  
 gttcttaaat gataggaatc tgttgtagt atgcctctgc aggtgtcaat tatctgaaat 1140  
 cccctcaatt tagtacgtat tatgagctaa caaatatttt tgttttacat cagtcttaat 1200  
 agtcc 1205

<210> 44  
 <211> 770  
 <212> DNA  
 <213> Homo sapien

<400> 44  
 acctcaaacc aagatctgtc agtgcagtca ctaaatatat caactggcca agttcacaca 60  
 atattccaga aacctacatt tctcattaag tagatagtat catggacttg agaacacatt 120  
 ttctggacca aataaatctg gaaaatgcta ttcttatgcc atcatattta agaacagtaa 180  
 tatatcattt taatagcttc tctgccatgt cacacatggg gaggacaaaa catcttttaa 240  
 ctaacaaaag agattcagag aggaaactaa aatcagagat cttggtagaa aaacatagca 300  
 aaaggatata aattatcgta aaagggttatt gttttattca aatagacaat ttacataaat 360  
 gaccagctgt atacttgaag ctagtggcaa cacatgtaaa agcctgcaaa accaattcca 420  
 gagatgaaat caattatgct gaagagccac aatactctta tagagcaatc aattatgctg 480  
 caaatattata ctttcaacta ctgactcaaa aaactgggac agtcctgtta catgagaaca 540  
 gctaattgta aaaaaagtga atctgatcac tttaatgatt ctacctatc ttatattaaa 600  
 ttttgataaa aaccactgaa aattggactt tttatgaaga aaagtgttcc ctgaacaaac 660  
 catgctttct cttttggccc tttggcctag caatgggtgct accacatccc ggacacatac 720  
 ttgccaatta tttctctcta ctctcaacct ttctactttg cctaaattgt 770

<210> 45  
 <211> 614  
 <212> DNA  
 <213> Homo sapien

<400> 45  
 ggcttggccg tgcaatggga aataaattag tgcctatgag taagttaaag tgagcctgtt 60  
 gtttaaaaac caatacttca caaaatgtgt atgctcttta tgtgtgtgtg tttggggcag 120  
 tccctccctt tagttcatag tagagttccc ttttgaaatgc tgccgcccac ctttagcttt 180  
 attaaggggg cacaaggaag gaggattttg tcttttccat acttgggcca aaatttttagg 240

tggaagaggt aaatcttttag ggacaaatat ggtcacggca ataggatgca aaggagagaa 300  
 gctggcaaga gagaaaagga aattcagaga accagattag gataaaaagtt ttcagttgag 360  
 ccctccgtct tggagagaaa atatctgaaa cttgtatttg ctcaaacgct ggaagccatt 420  
 ctatgcctgg gggagaggag aagagttcag tgttccagat gactggcaag gcttgaatgt 480  
 caagccttgc tgcaacacag acccggaac cttgggagtt cccatcggcg gtggtccaga 540  
 gaagataccg taatgtcaca ttacatctta ttgtcacctg ctctgtaaac cgaattgcc 600  
 gcacactggc gcc 614

<210> 46  
 <211> 656  
 <212> DNA  
 <213> Homo sapien

<400> 46  
 ttagcctggg cttcgcgccg aggtacactg gttctcgctt cagcgaataa gggggacaca 60  
 ctggagggtta acgtgttttg gggagaaagt tccaggtttc tgggtcataa ggtggaattc 120  
 aaaaaatttt cccgattgac gacttgattg aaagagtgtt attctataag ccctgaaatc 180  
 catagaatag gagcgtctgg gttaaagaat aaaggggttg tggagaccac aaagttcctt 240  
 aattaatgca gaatgaaagc ctccaaggta aacaaggaat tcagaagaga aataagaatt 300  
 gtaaaatggt ttcttgtcaa agaacttata aaaagctgtc agaaactctt aggttttaat 360  
 tcctcgtcct cgagtcacgg agtgaagatc ctggagaaag ggagaaagga gttctttcta 420  
 ttcagataat gtagacattt tttttcacat aagaagacag cctttgtaaa ggtcaatgtt 480  
 caaaaataat gtcacaagaa aatataaatt ttatgggttc aaaataacct tggagatttc 540  
 ttttaagggc ctgctattct gtcaagtgat gcttataact agagtttagc tggaaatttg 600  
 gtatcttttt tttttttttt tgaaacagag tcttgctctg tcaccaggg tggagt 656

<210> 47  
 <211> 550  
 <212> DNA  
 <213> Homo sapien

<400> 47  
 actaatttta ttctagcttc ataggaggt tgtatttgtt gtgtggtgtg tgtgtgtgta 60  
 gtgtgtgtgt gtgtgtgtgt gtgagagaga gaggetcatc tctgttgccc aactgctgga 120  
 gtgcagtgcg cccaactott actctcaact gtaaccctca actgctccgt ggacccaact 180  
 gactattcct tcctgtccat cacgtcctcc caatagtatg cttgagttat tatactgcta 240

24

gcacaaccaa ccatcattgt gtgctatatt ctttaacata ttgtgggtgt agtaataaga 300  
ggatcgtctc atctagtgtt cgcctagggt ctagcgtact aagaatctcg ttggatctca 360  
gatgtgactc ctaccattg agggcctccc aagatgtgcc agataattac atgtatgaga 420  
cgccagtgc cccagaccat aaccaacat ttttaacatg tgcatataac aactacctaa 480  
tatctaatat gtcacaattt tcaatctctt ttcttttaac caatttcaac ccagaaaact 540  
ccaaagaagt 550

<210> 48  
<211> 384  
<212> DNA  
<213> Homo sapien

<400> 48  
tttccttccc aaagtgcata tttttaagac tatccataaa atgcttccta gggcaagcat 60  
cctccaaagg gtcttggtta aggactatgg cagaccccaa gactgggtca taatctaggg 120  
tggaaccaagg gggctagccc aaaagggaag ctaagtgtta tgactagatt gaaactctgg 180  
tgccagctat tttaggtttt cacatacaat tctttatata actggtaaac cataaactgg 240  
cttccccctt ggtggatata cttttaagta tttctgggat gtgtttatat ggcagtttagc 300  
tgaaagtcag cagtcagcta aaatcttgta atcaaataat gcacaaggta gtgatattta 360  
cttgaaatag tagtacctgc ccaa 384

<210> 49  
<211> 327  
<212> DNA  
<213> Homo sapien

<400> 49  
acgtgaattt aagactaaat tttctttttg gccatgttta ttctttccat atattctgat 60  
ttataatgtc ttttattatt gtcattttat ccatattggg cactattggg ttttattttt 120  
atttgaatcc agtgattctg gtgattgatt ttttgtaatg aaagtattaa aataccagtt 180  
gataacatct tagatatttt ctttttgatt tttgtttcca gctctgttaa taatttctaa 240  
ttttgctcct attgtaaaca gagaatactg gccatgcaat tacttcattt ttttgcatt 300  
tattaaatat tcatttctaa ttgtagt 327

<210> 50  
<211> 485  
<212> DNA  
<213> Homo sapien

<400> 50

25

acccctccgg gggcgccctgg gaccctcacc caggccaggg ccttcgggga gtagcgtata 60  
 ggaccgggca aaactggggg gccctcctcc gcgcgggaag gacctggggg cgggattaca 120  
 tcagggatgg ggcgtggcca aaaccagagg aagggtggt gtgtggccac agttttgggg 180  
 atgggtgctg tgtctctcac caccctcct tttgctggtc aagaatgtat ctgcttttct 240  
 ggagcaaggc ccagaccttg ccgctttegc tgtgaattct ggcctctggg caggccccct 300  
 ggaggcagaa cctgcttttt tgggcactgt ttgctgaaca gggcacagat ggccatgtga 360  
 cctctgaaat tataaatata aagttaaaaa acaaaaaaaa ccttcactaa acacaaaagg 420  
 ctcccagaaa tcaaattttc cagggaatgt ggaagggtgt ccactcccgg gaaaatttta 480  
 atgtc 485

<210> 51  
 <211> 431  
 <212> DNA  
 <213> Homo sapien

<400> 51  
 actcattcta aagagggtag attgtatcca ttttaatagc tctcaatacc cttgcccatt 60  
 ccattacag ctggcagaaa ttcccagct gcaaatgggt tgcattttatc tgttaccat 120  
 tcttgatttg cagcacgatt agactgttct cttttctgcg tggcttatgt gtcactccat 180  
 gagaatctgg gagaggagaa ctcgatatgg tgctcattcc atcattttta ccatagtgtt 240  
 ctttattgtg aagagggtatc ccggttagag tggagtcaaa gctttatagc aagtgtattt 300  
 caggtaacta ttcttttact tgtttcagat ctatttcagt agcacagcgc accaagatag 360  
 aacacttata agcccttatt ttggtataga tcaccagaga aaccactcac cactcgcaca 420  
 tggccttcag t 431

<210> 52  
 <211> 605  
 <212> DNA  
 <213> Homo sapien

<400> 52  
 ctgggtacga cttggatcac tagtcgggc cccatgggct ggaattcgct ttaccgtgtc 60  
 ccgccgcgag gcacacaaat gggtcgcca tcatcgggaa ggacgttcag attgctttca 120  
 actttggctc ttatgaataa tgcttctatg aataatcaca tacaagtatt tttggggaaa 180  
 aaaaaagtta tttctcttga gtaaatatct aggagtagaa aaacaatttg tgatcaatat 240  
 ggaaattaaa attgtattct actagatata cactttgtga tcaatatgga aattaaaatt 300  
 gtattctact agaaaaacaa tttgtgatca atagggaaat taaaattgta tactactaat 360

attctatatt gatcacattc cttagcttag agcttgactt tagttgatga ttctaattta 420  
 gaaacaattt tggtttttga attgggtaat ttttaaggga ttaacaaggt ttgaaaatta 480  
 tacgttttat tggatttttt tttttttttg agacagagtc tcgctctgtt gcccaggcgg 540  
 gagaacctgc ccggggggggc ggccgaaagc cgaattctgg ggatatccac cacactgggg 600  
 gccgg 605

<210> 53  
 <211> 425  
 <212> DNA  
 <213> Homo sapien

<400> 53  
 acttcactaa taaatgtaat aacttgtaag aaaaaaattt ggtagttac aaggaaagcc 60  
 tgttcaattc tcctacattc attattatta aaatatatgc atttagtcct aatcaaagat 120  
 aatcagcatt cttccccaca tttatagtaa acataatttt tatataaaat atttgtaagt 180  
 attggcagca tgcacaagca gcatgtgctt tttgtcatac attctcacag ttggtaaatt 240  
 aaaatcaaga tagatctatg ggactctata tcattaagat tactcaaggt ctgaaaaaca 300  
 ccttaaacc cttggtttct ctttcagtga ttaagcatag tctttctaaa ttagcttgtg 360  
 taaatgcaat caatattttt caattcaaat gctattaaaa taatatctgt attcaaaaca 420  
 atggt 425

<210> 54  
 <211> 482  
 <212> DNA  
 <213> Homo sapien

<400> 54  
 acaggattct taaacgtagt gggtatttca caatggatct attccaacca gggtcaccca 60  
 atttatgaat atttccattt tccaaattgg tttgtcagcc actttctcta ggaaacaaat 120  
 tagggtttgc acaaacaatt tgcttaccga gcttactaat gtgtagaacc atttaccgg 180  
 tgcagtgcac atcaattgat ggcttaaagc cacaatcagg ggtgactgct tctctgacca 240  
 aaaacaaata aaggtaagaa tgtataataa atcctaata ttttttttgc cgacaatatc 300  
 cccacaacct cagaatggtc tgctgcagag aaccttggtt tctgtatcag actaatgtct 360  
 aaaaaaactg attctaaaaa tataggcttt tgcaagtcaa agatataaga taggaataaa 420  
 tttttttttt ttctttttgg agacaagttc tcaactctgt acccaggctg gagtgcagca 480  
 gt 482

<210> 55  
<211> 836  
<212> DNA  
<213> Homo sapien

<400> 55  
acaggattct taaacgtagt gggtatttca caatggatct attccaacca gggtcaccca 60  
atztatgaat atttccattt tccaaattgg tttgtcagcc actttctcta ggaaacaaat 120  
tagggtttgc acaaacaatt tgcttaccca gcttactaat gtgtagaacc atttaccocg 180  
tgcaagtgcac atcaattgat ggcctaaagc cacaatcagg ggtgactgct tctctgacca 240  
aaaacaaata aaggtaagaa tgtataataa atcctaatac tatttttttc cgacaatatc 300  
cccacaacct cagaatggtc tgctgcagag aaccttgggt tctgtatcag actaatgtct 360  
aaaaaaactg attctaaaaa tataggcttt tgcaagtcaa agatataaga taggaataaa 420  
tttttttttt ttcttttttg agacaagtgc tctctctgtc acccaggctg gaggcagac 480  
tgagacctgt ctcaaaagaa agtgaaaaca attcttacct tgtgggcctt aaaaaagcag 540  
gcagcaggct ggatttggca tgcaagccgg tttgctgacc tctgctctac acttgggttt 600  
tttgtctttt ttttccctt tttgtggaga aagggggctc gctgtattgc ctgagcagat 660  
ctcaaaactcc tgggctctag ctatctctg gcctctgctc cctaagtgtc gggattacag 720  
gtgacctctg cgctaaacag ttggtatgcc ccacacatc acagccttac atttccatac 780  
ttttgatcat gttgtcccct ggggtctagat caccctgagg gttccctggg cctctgc 836

<210> 56  
<211> 824  
<212> DNA  
<213> Homo sapien

<400> 56  
cgccagtgtg ctgccatgag ggatacgaag atcttaatat taagtcttgc aatccatgaa 60  
gcatgaattg tagttttcta ttatttgtct ctttgatttt ttctgcaata ttttatactt 120  
ttcagtttagt catattttct ataaatttaa agaaaagaaa ttaatagtat atgcactctc 180  
atttttacta ttttatctta tatcagtata tagtcttcat tacctattac tatatccaag 240  
tattttctatt atctattttg tctagaagaa ctttaacgtt tottgtagta gaagggttga 300  
gaattagatc agagtacctg gaagccaagt agaatagaag tatatcaaga agataggaag 360  
tagctacaac ctatgctaga tcggtagaag aataggagg aataaagaat tagaccctat 420  
agatttcaat aacttgtagt atagaagtac ttttctgata gaaaacaaat gattatttag 480  
tcaaaggaat tcgcaaaagg aaaattcagt atcagccata cctatttgga tctacatgga 540

tattctaaat attgaccaag aggttaattgt acagagtagg catagaagg tcatcacagt 600  
 agtagtagta taatagtaaa aaatgtaatg tatgtaacta cttgtataga gtaaggaaat 660  
 tatgggatga agtgaactgt agcccttaaa aatgaaaacg tagaactaca atgatgtgga 720  
 aagatgtgca tgacacatgg aaaaaacagt taaccgaaga gcatgtttta aacaattttc 780  
 attacatat atgcagtttt cagtctgtgt acctcgcca agcc 824

<210> 57  
 <211> 675  
 <212> DNA  
 <213> Homo sapien

<400> 57  
 tacggcgcag tgtgctgcaa ttccgattac caagaataga agaaaagagt ctccgtccca 60  
 ccccttgccc aaaggaagaa agagctaate acaggagatg ttccaagaat gcctaaagta 120  
 tcataatgaa tagttctact aatgtccact tgttcaatgg aaagtatgga cctcttagat 180  
 tcgcttccaa attcctccca aaatgactgt caatactctt gttattgaac ctggaaaaaa 240  
 aagttttctaa tatattattg atcacttate aattccccaa actgacgaca gcagttctta 300  
 gattgaactg ttaaaccttg ttcacatga ttatgaactg aatgattgta tgcagtttat 360  
 gggttttatg tctgcagtca ttccctcatt ttccataga aatgatataa acaatgatga 420  
 tgtaatttaa attttattca attatttatg gggttttatgt ctgcagtcac tccctcattt 480  
 tcccatagaa atgatataaa caatgatgat gtaatttaaa ttttattcaa tttactggat 540  
 tttaaatgtt ttctacatgg agaccatgaa gaggaactat gttcagagaa aatgtctaca 600  
 aagcaggacc atggccaacc acttttcac taaaccaatt cactaaaagt acctcgcacg 660  
 cgaccacgct aagcc 675

<210> 58  
 <211> 596  
 <212> DNA  
 <213> Homo sapien

<400> 58  
 gtacaagctt tttttttttt tttttttttt ttggggaaaa ctccgggggg cccgggggtt 60  
 gggttttaaac ccccttgggg ggggtcagtc ccttcccttt gggctgcaag atttaatgat 120  
 ataggggggc ctccagagtg gggggagagg gggggggcg tctccattat atgccccca 180  
 ggtgttagga gaggggtcct ccatacaata agagattttc cggtttcaga aggagaagcg 240  
 ctcccaaat ccgtggaaaa tttttaaaat atacgcgggg ggtgagaaaa atgtgaggtg 300



29

aacccttacg agagtgaggg gaatatccac gagggggggg tagggcactg cgggggatag 360  
cccgagaaaa gggcgggggg acatccggga aatagccgga aaaacgtggt gggggcgaaa 420  
acgggttaaa tttaacgcc ggagaaaata tagtatatgg gaacggggat gttgcggcgc 480  
aagggttggg cccaatggg tgttccccct gaagaatgtg gggaaccccc gggaagatga 540  
aaggcgccca tattaggggg ggaaaaacag cgccccaaat ggtagggagc attctg 596

<210> 59  
<211> 813  
<212> DNA  
<213> Homo sapien

<400> 59  
acttacttgt taagaaatcc atacattgta tataacttta gaatttgtct tcctctctgt 60  
tttattaaac ccatttatct tttgtgagag atcataaaca cctaaccag gaagggactg 120  
gggaacaagc aggcactgta tggagggcag acaggcccaa acccaggtct tccgtctctg 180  
caggagcac aatgtgtgca aacatatcaa gaaaagtga cattgttaca gacactgcc 240  
gaggtaaagga gaaaaaaatc aacatctggt aaaagccatc ccaaagcttt gcacacacac 300  
caaaaaaaaaa ggttgattgg tggaaatgta gctactaata ataaactggg ctccctaatta 360  
acaggatata actatggcta aggataaagc tgaattgagg cgtatatatt actgatgaag 420  
tatttgtgtg gtttgctagt tgtctccatg catgattatt gctgacctat ctccaggacag 480  
cacatatgac ttcctaagaa taccactacc tagcctactc attcagtgga tgtgacatga 540  
agtttccagg accagtagaa ttataatggg atatgaatat aatcttcgga gctctgtttc 600  
gatgaagtat ttgggtagtt cttagaagaa tattctaaat atgtcacatt catgcctagt 660  
ttttcagtg cccaatttgt gaattcttaa agagggcac ttgtcgtagt ctgttagcgt 720  
tgctgcaaac gcactacctg aggtgagta attggtaaag gagagagggt tacttggtg 780  
cacagctctg cagcctgtaa cccgatgggc aga 813

<210> 60  
<211> 1220  
<212> DNA  
<213> Homo sapien

<400> 60  
acttacttgt taagaaatcc atacattgta tataacttta gaatttgtct tcctctctgt 60  
tttattaaac ccatttatct tttgtgagag atcataaaca cctaaccag gaagggactg 120  
gggaacaagc aggcactgta tggagggcag acaggcccaa acccaggtct tccgtctctg 180  
caggagcac aatgtgtgca aacatatcaa gaaaagtga cattgttaca gacactgcc 240

gaggtaagga gaaaaaaatc aacatctggg aaaagccatc ccaaagcttt gcacacacac 300  
 caaaaaaaaaa ggttgattgg tggaaatgta gctactaata ataaactggg ctccctaatta 360  
 acaggatatc actatggcta aggataaagc tgaattgagg cgtatatatt actgatgaag 420  
 tatttgtgtg gtttgctagt tgtctccatg catgattatt gctgacctat ctccaggacag 480  
 cacatatgac ttcctaagaa taccactacc tagcctactc attcagtggg tgtgacatga 540  
 agtttccagg accagtagaa ttataatggg atatgaatat aatcttcgga gctctgtttc 600  
 gatgaagtat ttgggtagtt cttagaagaa tattctaaat atgtcacatt catgcctagt 660  
 ttttcagtgc cccaatttgt gaattcttaa agagggcatc ttgtcgtagt ctggttagcgt 720  
 tgctgcaaac gcactacctg aggctgagta atttgtaaag gaaagaggtt tatttggctc 780  
 acatttctgc ctggcagtgg ctatagttag cttctagtga ggcttcagga agcttttact 840  
 catggtagaa ggcaaggggg ggcgaggca tgtcacatag caagaggagg agtgagagag 900  
 agaaggaggt gccaggctcc tttaaattgt aaactaacag caagaactca ctcatccca 960  
 tgaggaaggg accaggccat tcatgaggga tcctccctca tgacccaaac accccgacta 1020  
 ggccctacct ccaattctga ggatcacatt ccaacataag attcagagag tgaaaacatt 1080  
 caaagtctac caggaggtga ttaaagtgat atgacaatgt gaggtatatg ttacctggaa 1140  
 ggtagtagga gcagatggg ggaagaagac aagtgtagca actaccaggt cctctgcatg 1200  
 aaagagccct cctcaaagag 1220

<210> 61  
 <211> 347  
 <212> DNA  
 <213> Homo sapien

<400> 61  
 tcggatcatt ttttaacaga ttcattggagg ggtattggat atacaataac cggcacatat 60  
 cgaaagtata taatttaagg ttttatatta tggatatacac tccatggaaa ccacttaaaa 120  
 ttggggaata tatccatcac tactcccca aaattttcct catgaattcc tttgtaattt 180  
 cattaccttt cttcccatc tccaggaccc ttgcatcctc aggtaccac ggatctgcct 240  
 tctctctata tagatagggt tgcattttct agaattttat ataaatggaa ttacagatta 300  
 tgtgcccttt tcttgtctta gttctttact cacaataatt cgggggaa 347

<210> 62  
 <211> 470  
 <212> DNA  
 <213> Homo sapien

```

<400> 62
gagcgcagtg tgctgcaatt cgtattgggg caggtgaagc accatctggg tttgatagca      60
ccgagtatat taactcatct gttaatTTTT gacacatcat tgatttattt atttagaaaa      120
attccatcct gttttgctgt gttgatgttt gccattatga aaaaaattgg atggcatttt      180
ttaaaaaaga catagtccat tgatttaaga caagggtggg gcttccaaag atgctttgat      240
tttctaactt ctattttcta tgccatgaag ccctctttta aaaataagtt aaattaatgt      300
tttaaggaaa tcatctattt tgaactagtt agaaatcggt gggtagatat aacatatgaa      360
tacatgatag taggtcacta aaaaaaattt ctgcttactg catttttaat cattaatata      420
tttcaccctt tcgctgggct aaaatcaaaa gggtaccact aatcttgaca      470

```

```

<210> 63
<211> 688
<212> DNA
<213> Homo sapien

```

```

<400> 63
acccccttcc actccttaga aaagtTTTTT aaaaatattt ttctacattt atttattttt      60
tccctaaatg ttgatgatgt ctttttttct acattcaact gcatctctgc tgctttattt      120
atgtatttag gcctaataca atgactgggt tatcttttaa aactcgctgg tttccaaaaa      180
atataggtat tcccctttta atttcattgt tcttttcccc aaattttgat acaacatttt      240
agggggaaaa taaaaagggt gtgataaaat gggaaaaaga gaatgccctt gctcagaaaa      300
ataatttttag gggaaaaata aaataagggt gataatgaaa gaaagagatg caccttgctc      360
cagacattgt tattaagcgc ctattttaag gttttctcta atattttccc tottaccctc      420
tagtgtgtgg ccccatgggg acccacgcaa acccggtaaa aaaaaccgaa acctagccct      480
gttttaattt ttcttttttt tttttttttt ttgagacgga gtcttactct gtcaccagg      540
ctggagtgca gtggtcgcaa ttccggtcca cattttttga cctgccaaag cgaattccca      600
gcaaattggg ggcgaaatat ggtccgactc gtcacagctg gggtaccaga caaagcgtcc      660
tgttgaatgt tccgtcatca aatagtca      688

```

```

<210> 64
<211> 807
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (125)..(125)
<223> a, c, g or t

```

<220>  
<221> misc\_feature  
<222> (133)..(133)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (135)..(135)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (137)..(137)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (139)..(140)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (142)..(142)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (145)..(145)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (147)..(147)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (150)..(150)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (155)..(155)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (159)..(159)

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (163)..(163)

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (165)..(166)

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (178)..(178)

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (247)..(247)

<223> a, c, g or t

<220>

<221> misc\_feature

<222> (358)..(358)

<223> a, c, g or t

<400> 64	
actgacaatc tcataaaatc tgtttttatc agattccctt cccacactgg agaattctca	60
tttggttggc ttaagcttct aaagcctgga tgattctata tttattaacg gaaccaactg	120
attcnctttc aancnangnn anttntnggn gtggncttnt ttnannccca accccaanga	180
aaaattctag ttttcttttg cctatagggt tttatattgt ttgaggcaac aagcattacc	240
ttttttngat ttaccttttg aggtgacaac taaaggaaac aacacaccct cctaattttc	300
aaattttctc atctctgcat gtgattatct cttttcctaa tatgctactt aggttttntg	360
gaaacaaaag acaagaattc atagtaattc ctttcaaata gcagtttctg aagatttagc	420
ccattcatca aatctctatg tatcatttga ttctgctttt ccagtgaatt tttgtcatat	480
caatgaacct tatcatctgc aaatgtatta agtatactct ctaattcttc attacaggca	540
ttaataaaat ggcaacagat tggaatcaac atattaatag cttctcggtg tgataatata	600
agcactttct ttccacctta attgcattag catctaggcc atattttctt gtattattcc	660
taaggatatc ttgagacaca ttgtcaaata ttttggccaa tgccagacct aaaggggaaa	720
taagagaatt caggatcata tcaaacaatt aataatttag ttgagtaata ttccctacaa	780

atattaaaaat aattgaaatt gtatagt

807

<210> 65  
<211> 257  
<212> DNA  
<213> Homo sapien

<400> 65  
gcttcttata cagcctgtca gataccatga tgaccaaaca atcctcttgt gcttgtaagt 60  
ttcccagcct caggtattcc ttaccagga acacotagaa cgccctggg atcggttaact 120  
aagtgaacta atatcgttat catatogtca attcaagcat gttaaaaaa gctaaacgaa 180  
caaaactacc aagagtgaag gacacagcgt cagtaccaga actccagcca tttagaggca 240  
tcttctcaca atagggg 257

<210> 66  
<211> 898  
<212> DNA  
<213> Homo sapien

<400> 66  
ggcagtgtct gcagtgccgt attggccgag gtggaactta tgcgtttttg ctgttgatcat 60  
ttttcaactg tgacccttgg attggttggt tggtgggta atgaattcct tcagaattat 120  
gaaggcattg ctacatgggc ttctagcttc ctgacactgc tgtggagaat gagaagtctg 180  
aagcctttta attccttggt ctttctaggt gacttttccc ctgctctgaa ctgtcttggt 240  
ttccaatggt ccgaaaattg caagtgaaga tgttcttggg tgattggcta ttagtcttcc 300  
cattgtgctt gggcaactca gtgcccctca cttttgggga ataactctgt aatcatccta 360  
ttggtgattt cttcctatct tctctattcg tttccttttt ggaactccta ttacttggac 420  
aatggacctg ttgaactggc ccttctaatt tcttattttt ctctctagct ttgcatcttt 480  
ttgtacttta ttgttctggg agatttcttc aactttatgc ttcccattct tctgagtgtt 540  
ttccatcttt taaaggtttt aatcactttc actttatttt tattctcoga attgtccttt 600  
ttatagcata ttttgtttca gggttatctt atctcagaag ctgttactga tagcctctgt 660  
catttttttt tgtctctaca taatctctgc tccaaatttc ttttcttttt gttttggtct 720  
ctgtcttcta tattagattc ttttcttggg ggtcttagac tccttggttg tctgctcaca 780  
cttaaggatg ggtcaccta aagctgactg gaagcatttg ttctctgttc aggtgggtacc 840  
tgcccaagcc tgaattctgc aagatactac attaacta gcaggccgca cgacgcat 898

<210> 67

<211> 677  
<212> DNA  
<213> Homo sapien

<400> 67  
tagtcggctc gcctccactt gtaacgcgcg cagttgtgct ggtaattcgt gcttaccatt 60  
ctcacaccag ttggaaatgg ttattattaa aatagtcaaa ttaatatcat gctggtggcc 120  
gggcgcggtg cctcatgcct gtatcccagc actttgcgat gctgaggcag gcatgatcac 180  
aatggtcagg atgatcggag accatcctgt gcctaccact agtgacaacc ccgttcttct 240  
actaaataat aaaaaaaaaa aattagctgg gtcattgggtg gtgggcatcc ttgttagttc 300  
ccatgcttac ccgaggagtg ctgaggcagt gatttactag tcattgaatc cttgggatgg 360  
ttgagatgct ttgcaggtga agccaatgat tagctgccta cttgctcctt ccaggtcctt 420  
gggttgacag gagcgcgtac ttcttgttct tcaacaaaca acaacaacaa acacaaacac 480  
acacccaaaa agataattac acatattgag tgtggacaaa aaggcattag aagagagaga 540  
aacacagaag agacacacac atcttattta ttgtgtgtgt atatgggagt gggagggtgtg 600  
tctaaagtgt tagtcatcca gaggccaagt ctcttgggaa aacacgacca gagtgggtgtg 660  
ggtgtgcatc ttctttt 677

<210> 68  
<211> 3809  
<212> DNA  
<213> Homo sapien

<400> 68  
aaaggaagat gcacaccac accactctgg tcgtgttttc ccaagagact tggcctctgg 60  
atgactaaca ctttagacac acctccact cccatataca cacacaataa ataagatgtg 120  
tgtgtctctt ctgtgtttct ctctcttcta atgccttttt gtccacactc aatatgtgta 180  
attatctttt tgggtgtgtg tttgtgtttg ttgttgtgt ttgttgaaga acaagaagta 240  
ctcgctcctg tcaacccaag gacctggaag gagcaagtag gcagctaate attggcttca 300  
cctgcaaagc atctcaacca toccaaggat tcaatgacta gtaaactact gcctcagcac 360  
tctcgggta agcatgggaa ctaacaagga tgcccaccac caatgacca gctaattttt 420  
tttttgtatt atttagtaga agaacggggt tgtcactagt ggtaggcaca ggatgggtctc 480  
cgatcatcct gaccattgtg atcatgcctg cctcagcatc cccaagtgt gggattacag 540  
gcatgcacca ccacacctgg ctaatttttg tatttttagt agagatgggg tttctccatg 600  
ttggtcaggc tgggtottgaa ctctgacct caggaaccct gcttgggtgtg cctgctgtaa 660  
accagtggtc gggatgggcc cgaggcggcg ctgagagagc ggccacgatg gccgcagtcc 720

gcgggtgtgga ctctcttgca gccagcagcg cgtgggatgt gctgtgctcc cagagaggat 780  
tcagggcact aggaaggagg ccctccctgt gcctggagca ggagggagca cttcaaaaag 840  
gaaatggctt tgaaggagga gaaagtcaga aggaagatgt ctcaggaaag caggaacatt 900  
tgaggagaag gagagcacca ggtgccccag gggtagctag ggatgaagct ggagaggctc 960  
atgccaggtt tagccccttg aatgtgaatg ctaaaaacct gttggatttt acagcatctg 1020  
ggagctctat catttgcttt tccaagtctc ccaccaatcc aaatgatccc ccagggtgtg 1080  
tgtgtgtatg tctaggaaat gcagacctca atgacattgg agtccgagaa gccagggagg 1140  
cgctggctct tgcagatgac ccagcatcc tcatcgtgcg tacagtcact gttccccag 1200  
ccccgggagg cacattcagt cacactctgc tcgggtccac tgcagctcaa gttgtccagc 1260  
cagatgcccc ctgtggaagg ggagatgaag ggacagagaa gcttggggag ggaggatggg 1320  
ggagtaatgg ggatgtaggg aagaaggatg tgggcagcct gcagcgtgaa gtcacatcgc 1380  
cagatgggtc ccatttcacc agctcgtgt atctccacgc ggccctcgta gggcttctg 1440  
gggaagccag ccagccggaa ccgaagcccc tggtccccg ccttcttctc agggcccgtg 1500  
gaaggggacg gagaccccaa gcacgaactg cacagcaggc acagcagcag cccccagggg 1560  
ctccactgcc agacactgac aggtcgcagt gcagggaagg cctgggtgcc ccagagacaa 1620  
agtggccagg aagcgcgaa ggaaccgccg ggggccatgg acggagcagt gatggaaggg 1680  
ccgctttttt tgcagagtca gcgctttggg accaaggtag tctggcgcat ggatgccgaa 1740  
ccttatccgg gctagtagtg ggtgagagag ccagaaaaca gggagagagg tggaggaaga 1800  
cctgggcgt gctctaccg gccagtcccc acggcgtagc gcggctcgag ttctttgacc 1860  
ataaggggtc gagctctggg ggtggccgag ggagctcgcg ccgcctggac tgcaaagtga 1920  
tccgtctggc tgagtgtgtg agtgtggccc ccgtcacctg ggagaccccc cctgagcccc 1980  
gcgccactgc cttccgcctg gacactgctc agcgtcgca cctgctggcg gccgacgcgc 2040  
cgtccagtgc agcctgggtg cagacgctgt gccgaaacgc ctttccgaaa ggcagctgga 2100  
ctctggcgcc taccgataac ccacctaagc tttctgccct ggagatgctg gagaactcct 2160  
tgtacagccc tacctgggaa ggtagacgcc tcagaagccc gggcagggat ggagtgaaga 2220  
ggaggagggc cgagggcctt tgggaagtgg gtggataccc aggggccccat ggggaagtaa 2280  
gaagtaggaa ggcctgaga tctggcttcc gactgtctaa tcgtgtatgc cttccaggat 2340  
cccaattctg ggtaacggtg cagaggactg aggcgcgca gcgctgtggc ctgcatggct 2400  
cctacgtgct gagggtggag gctgaaaggc tgactctcct gaccgtgggg gccagagtc 2460



37

agatactgga gccactcctg tcctggccct acactctgtt gcgtcgctat ggccgggaca 2520  
aggatcatgtt ctctttcgag gccggccgcc gctgcccctc aggccttga accttcacct 2580  
tccagacggc acagggaaat gacatcttcc aggcagttga gactgccatc caccggcaga 2640  
aggcccaggg aaaggccgga caggggcacg atgttctcag agctgactcc catgaagggg 2700  
agggtggcaga ggggaagttg ccttccccac ctggccccca agagctcctc gacagtcccc 2760  
cagccctgta tgctgagccc ttagactccc tgcgcattgc tccatgccct tcccaggact 2820  
ccctatactc agaccccttg gacagcacgt ctgctcaggc aggagaggga gtacaacgga 2880  
agaaacctct ctattgggac ttgtatgagc atgcgcagca gcagttgctg aaggccaagc 2940  
tgacagaccc caaagaggat cccatctatg atgaacctga gggcctggcc ccagtccctc 3000  
cccagggcct ttatgatctg cctcgggagc ccaaggatgc atgggtggtg caagctcggg 3060  
tgaaggagga gggctatgag ctcccctaca accctgccac tgatgactac gctgtgccac 3120  
cccctcggag cacaaagccc ctccctgtct ccaagcccca gggccagcc ttccctgaac 3180  
ctgggtactgc aactggcagt ggcatacaaaa gccacaactc agccctgtac agccagagaa 3240  
tccagatccc aggaagagga aaaggagaag gaggagaaac aggtggagga ggaggggagg 3300  
aggaggagga ggaggaggag gagaaggaga aggaggagga ggagaaggag gaggaggaga 3360  
aggaggagga ggagaaggag gaggagaagg aggaggagga gaaggaggag gaggagaagg 3420  
aggaggagga gaaggaggag gaggaggaga aggaggagga ggagaaggag gaggaggagg 3480  
aggaggagga ggagaagaag aaaaagaagg aggtaaaaat ggctttcaga gttgtttgct 3540  
cctctgcac tgcactatc acatctcatc ttgaagccct tcctattttg ggcacctaca 3600  
ttggctgctt tttctcaaca aacataaatg agcgcaaaca caggtataca catatgcaca 3660  
cgtgcaatgc agacagaccc ctaagaacac accagcacac agatgtgcac aggcgcactc 3720  
acacccactg ccaccaatth agagacccaa ccacagacac actcgggcac aaaaacacac 3780  
aaagacacga aaaccatac aaaggctga 3809

<210> 69  
<211> 485  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (461)..(480)  
<223> a, c, g or t

<400> 69

38

acagtagcga gattacagct cactgcgagt cctctaccct tccagggtca agcaatcctc 60  
ctaactcagc ctccaccacg ttcagcctcc agagtagctg gagactacag actgtgtgcc 120  
tggcttttgt gtttgtgtgt atgtatgtgc tgatgtgtgt gtgtgtgtgt gtgtgtgtag 180  
acacaggggtt attgtcatag attcgccccg tgetgatctc aaatctagtg tgccctgaaag 240  
gggccctcct atctcagctc tcccagagct gctgggttaa agtgtgtgca gccactgtgc 300  
ccagtctcag ttgtgctctt ttgaagaata aatatattcg ttattctgcc atattgcttt 360  
ggaagagagg acaacagaga tatatagggg caacgcagaa taaatcctcc tgattagttt 420  
agcgctcttt tctagtaata tatacttggc cattttatgt nnnnnnnnnn nnnnnnnnnn 480  
aagcc 485

<210> 70  
<211> 580  
<212> DNA  
<213> Homo sapien

<400> 70  
cacaatcaat actctttgaa ctaatagagc tttaaattca gggtaattcc tcaaaatctg 60  
gaagcttttt caagaaaact tttcaaaagt taagaacgta tgacccacag gcctgtgtgg 120  
ctcaacgcct gtaacccacg aaacttcttg gaaggccaca aggttgggca cagattcaca 180  
caaagggtgtc acaggagagt gttcgggatt gaagctctcg gcgcacaaca cttgggtgtga 240  
gaaacacctc gtgtcttctc aactataaaa taacacacaa aaatatacgc gaagtgtggg 300  
ggctgggcac acttgtaagt cccacagcta tactcacagg aagcgtgtga aggcacagga 360  
gagaaatggc gtgtaaacct gggagaggca cagagagctt gcaagtgaag cccacagagt 420  
tgcaaccacc tgcacactcc acagcctggg cgacagagca aaaatcgggt tcaaaaaaaaa 480  
aaaaaaaaagaa attgcttgta ttatgactgt gtattatgac tgtgtcaatg tcacaacttt 540  
tctttgtgaa tattgtacct gcccgggcgc cgctcgaaaa 580

<210> 71  
<211> 715  
<212> DNA  
<213> Homo sapien

<400> 71  
gggtaccgag actcgggatt tcacttagtt tacggcccggt cacgtgtgct ggaatatcgt 60  
gctctttgaa gcaggccgct cgggggcaag gtaaacaaag gttaaattat gtgaccgacc 120  
aaaatatatg tgactgcaca tggttctata acaaatacgc aaagaattcg cccgcgctta 180  
tattctagtg taggggtgaa cattacacaa caatattaca acattccttg gaataataat 240

attgataaaa atattgaaca atatgttgcg cagtgtgtga agaaaatata gagcatttgt 300  
agagagattc cagagtattg tgggtgtgtg aagagtatgt ttgggtgtgt tctctgacaa 360  
tagttcttgg aggaacgacg gagaagcgtc taagcgtcta tgcccttctt tctgcattag 420  
ttttggcgtc tccacacaag cgcgtctccc acacaaatct aatcaacttt ccgtgccgcc 480  
ccattataag ggtatagaag gtgaagcaat gtctccgtca acaaagttgc gaacaatgtg 540  
cacaagccat atcacgcaaa ggccggcgaat ctacaacaaa aaggcgactc aaagaagtgg 600  
aaatttttta aacccaaagg aacgaagaaa aaacaacttt caaaaaaat aaaaagaaaa 660  
accaaacccat attttgccac atgtgagagt acctcggccg gaccacgcta aagcc 715

<210> 72  
<211> 324  
<212> DNA  
<213> Homo sapien

<400> 72  
cacgaaaaaa aaacccgcac gcacacacca ccacgacagg acgcccctca ttaatacgcg 60  
cgcagtgtgc tggacatcgg cttacaagga ataagttttg gtgatctact gcatagcaaa 120  
gatgactata gttaacgaca ctatattgga tttttttcca aattaccagg aaggagagaa 180  
ttttgaatgt tctcactata aatgaaaaat ctttgaggta ataaatacct tgattttaatc 240  
atcatacaac atatacatgt atcagaacat cacactgtaa gcctgtatta tcgcacgagt 300  
acacttaata cgatggcgga cgcg 324

<210> 73  
<211> 751  
<212> DNA  
<213> Homo sapien

<400> 73  
agatagatat tccctattta tatacctcac attagaagta aagagctgtg atactaggca 60  
taaaacgaca caatgcttac acacaataat attgctgtga taatacactt ttctgaagtc 120  
aaatttctga ttatccccta gcagtctagc agaacaaagt ctgccacagt ctctcagagc 180  
aggcaaccat ttgctgtctc cagatccgt agctagattg ggtaggtagc ccatacaaca 240  
tatgagggca gagcgatacc catctagaat ccactcagat tcacacacta gtactcctca 300  
ccctaaccac caaaataatg ctttaccagg tttctaggta ttccttaatc tagtcaagtt 360  
gacattttaa attactgata cctaaaatga agtccacaag tatcatccct tatcaatctg 420  
gcatccatac acagttgtat gtaaaaaaaaa aaaaataaaa tacaagaaaa gggaacgcac 480

40

attaggatcg caaggagaca agagatagca gagggaacca aaacaggagg acacagagga 540  
acgacatacc ccagaggggg tcgacagaca attattacac catggacggg aaaccacaca 600  
aaacaagata gtagagaata ataaagaaca gatgaggccg aagacgaggc aaataaagag 660  
gccaaagacg caaaagagga gagataaaac agtcgcggaa cacacaaagg atgaagaccc 720  
ggagacccaa taagaggaga caccattga c 751

<210> 74  
<211> 186  
<212> DNA  
<213> Homo sapien

<400> 74  
attoggcttt cgagcgcccc cgggcaagt gcctataatc ttagcttctc tggagccctg 60  
agggtgggagg atctcttgaa attggtgagg tcgaggggtgc aatgagccat gattggacca 120  
atggaattcc agccgggtca tcagagagac actcaatctc tctaaaaaca aacaaacaaa 180  
caaaca 186

<210> 75  
<211> 569  
<212> DNA  
<213> Homo sapien

<400> 75  
tcgagcttgg atccattatt acgcccccaa ttgtgtgaat tccgcttagc gtgttccgct 60  
cgaagtactc ttccaggcag aatccttcca tgtagactag aagaattaca tgaacacaag 120  
cctagactgc aagaggagac aggggcaatg tagggagcac tgtaaaaaaa cacatcgaca 180  
tcccccttgg ccacctctga tctcagccgt ggctacaagt ggacaagttg cgcctattgg 240  
agctcagatg ctgctctcac tgacagcgat cctcatcgtg catgaggtag cctgttcttc 300  
agcgtttcca ccccaggctc gttctccagc acccatggag caccacaagt ctgtctaagt 360  
ctattttgtg ctgctatgag agaatgctac agattggcaa gcaatggaag ttttgggggg 420  
ctcatggttc tggagactgg gaagtccaag atcaaagtac cacatctggt aagggccttc 480  
ttgctgcatt gtaacatgac agaaggcatc acatgggtggc agaaggcatg caagagagag 540  
agaaaaggag tttgagctca ccctttata 569

<210> 76  
<211> 255  
<212> DNA  
<213> Homo sapien

<400> 76

41

cataagacaa tatgaggggg tggctctctc ccttagatgg agaggtatag gtgtttctag 60  
 ttggtaaagg gtgggaatga agaaaagttc ccatgtgcaa gttgaagatt tgcaaagttt 120  
 aaattttctca cagatgctag aggaggaggc atacatgggc tttcttgtca gcctgtccag 180  
 atgtggacca aagaaaaatg ggagtaaaag tttcaaaggt acctggccgg gcggggcgctc 240  
 gaagcgaat tccag 255

<210> 77  
 <211> 1016  
 <212> DNA  
 <213> Homo sapien

<400> 77  
 accacttctc atctgatact ttatatttgc taggctatct cctttgctta tttccttcat 60  
 gacgtttcac tacataattt ttcctttacc cttcaatgcc tccgttattc gtccactact 120  
 ctgactttta tgacaaaaat aattactgag ttacctatcc tcttccatat tttctacag 180  
 tttgtcatat ggtattatat gtgacactta tcaaagccag cgttatcatg ttatcatatc 240  
 tgcgttgcaa tagtaggtag tctgtctaaa ttgagattga caagaaagaa cacatttgtc 300  
 atattgatct cttgactgac ttggagaatt ccactgagcc cagttttctc ctgggtattat 360  
 cttaatattt ctacacgata ttagtgatgt cttaaatttt gtttactgta ttcgatatgc 420  
 cttattgaca ttattactta aattttcatt catcatattt gatttctcac taatatttat 480  
 tacaaatccg atgtgcatgg ttactggcgt tatgcccatt catgaatttt ggctcctccc 540  
 cttcaatgag agtgaagttg gaaagatgaa acctttaaaa gaaaagatag tataataggg 600  
 aagtattgaa tgacaaggca ttctttggca tcagaatgga gtggcagctt attacaatga 660  
 gaagttcaga agcaacttca cagaggatgg aaaaccacta aaatccgggc agataaatag 720  
 aaactatcta gtttgaaggg gattagatgt ggatgggttg gcacttttaa agccagatta 780  
 tacagtttag tttactgaaa ttcagagtga atgagccaaa tgattctgtg agaaacctag 840  
 acaggcatca gatcaagatg ccttccttct gagactgtgg ggaagctctc aaatatttta 900  
 tgcagaaaag gatcaaactc caatttcaga gatacaacat taaccaacaa ggttgaatgg 960  
 actggatgga gagatgaaca ggaacagagg taatgagacc actttacagg ataagt 1016

<210> 78  
 <211> 392  
 <212> DNA  
 <213> Homo sapien

<400> 78  
 ggcgcagtgt gctgacgtcg ggttacgtgg atgcggccga ggtacaataa tctgtttgca 60

ttgttctttc tgccctctaaa tttctctgtc gtaagtctag caatgtttct agaaaagcga 120  
tcttaaaaga tatatctgat tatgttattt gccttattaa aaaccttcag tggctcccaa 180  
tagcatacat cttagagtta ttagatgcc aagtgtttta tatcaaagca gaccacttta 240  
acctccaaaa ttttttaaaa atcagaatta ctactaacca aattttgtc ttttattaac 300  
cctgtaaaaa tatagatacc catgagcttt gaaatttcta tgtctagtag gctgctctct 360  
catgcacca tatatttctt tttgttttgt tt 392

<210> 79  
<211> 822  
<212> DNA  
<213> Homo sapien

<400> 79  
acaagctttt tttttttttt tttttttttt ttttttgta ataatggcca gtttattcct 60  
ttgggtctaaa aaatccaaag gtaattgggg gtttcacatt ttctttttaa agttctttt 120  
ctcattcaaa agtttttaggt tgccccatt gtgccctctc aataaccgtg gtttgcaaat 180  
ttatttgtga cttcttatat tttgtataaa caaaaactcc tttaattctc tttttgtgtg 240  
ttggtagaaa cacgttggtc tcattaaaaa gagtctacct ctggcgcgcg caaacacagc 300  
gccataagcc gcgaaattct ccacgcacac atgtgtggcg cgggccgtgt ataccataag 360  
tgtggaatct cacgaggcct tctcgtgtat caccacagag ccttgtgggg cgtataaatc 420  
tcatgtgtgc tacaatagac gttgttatct cccgtgtgtg tgtgaacaat gtgtgtgttt 480  
aaatcgccgg cgactcaca acaatatgtc ccgccacaca cacaatataa tattcagcgg 540  
gggaaacaaa aaaagggggg aggaagcgag tggagggagg gcagagaggg gcagagcgcg 600  
aggcgggagg ggggggagag gggaggagga gggcggggga ggagcgcggg gcggatggga 660  
cgacgggggc gcggggggcc gccggcgggc ggcgcgggcg ggggaagcga ccgggggcgg 720  
agggggagag gaagtgagag cgggtgggag gcgggcagag ggagagaagg atgataggag 780  
ggaggggaca gggagggagg ggggggagaa ggcagcgggc aa 822

<210> 80  
<211> 513  
<212> DNA  
<213> Homo sapien

<400> 80  
acattcaatc tgattaaaat acggacttta gtttaaaaat aaagtattaa ttttagtcca 60  
ttaatagtga caagtatatc atcctaattg aagatgttaa ttgtagagga aactgggtat 120

43

aggctatata taggaatggt ctagtatcat cacaattggt ttatgaatat aaagctatta 180  
 atgttattct aaaaaataca agtatgggtg atgattttat atttgtagct agttgtaaaa 240  
 ttaagatcta gtccaatggg aacatcttct catacaaaga agatgacagc atattgtgga 300  
 gcatggccca gactagctga ctgtggagag gatgtataaa agttatggtg acagtgggtca 360  
 ggaaggagta aggataggta agagaacaga gtaacatgat tttgtgttac tttatatagc 420  
 tgctgtaaaa gtttcaatat catcacctcg ccgcgaccac gtaatccgaa attccagcac 480  
 actgcgccgt atcatgatgg aactcggtcca tgg 513

<210> 81  
 <211> 141  
 <212> DNA  
 <213> Homo sapien

<400> 81  
 cacctgtaat ttacaataag gagtgcattt aattgggttc atcagtagtg tggacattga 60  
 ccaagaaaag aatcagaaaa cttgaaatag gtccataaaa tccccaaaaa aaagcaagac 120  
 ccaaaaaaaaa aaaaaatttg g 141

<210> 82  
 <211> 631  
 <212> DNA  
 <213> Homo sapien

<400> 82  
 agaaaaacca cagcaagagt taagggttt tcaaagagtt actttagaca atatatctgc 60  
 aaaagtgaca agaggaacag gaagaaggtc atggagtga tgggacaaaa acagaaagca 120  
 gagacgttac gattacacaa aaactaccta tgactaggac tggtaaaata gtctatggaa 180  
 tgatgtctag gaatatactt tgtaaatgat aaagcacgct tgaaatattt attatcatta 240  
 ctctacactg ggagaaataa tggaatttag agaaatgatt ttggaggtct aaatccaagg 300  
 cctaagagaa taaaaatttg aaagagaaat aaagaagttt aaagtaatat cataaaactg 360  
 ctataatgac aaggcataag tgataaaaaa taggaaagaa aattataaaa atcaaaagtt 420  
 aaaaactgct caatatatta ggaagtataa aggtatttat tactttaagt taaaaaagat 480  
 aatatacctc tottaggatg aatgaggaaa atatatggtg gtcacgtaac aaggttaaca 540  
 aataatttgt actgcccggg cggcgctcga aagccgaatt ccagcacact tcgggcgtta 600  
 tagtggatcg agctcggtac aagctggcgt a 631

<210> 83  
 <211> 486

<212> DNA  
<213> Homo sapien

<400> 83  
acatctgctg gcaacaaact ctcttagttt tccttaatct aagaatgcct tcatttcttc 60  
tttaattcct gacagatatt gtcattgggt agaagattcg gagttgacaa ttctttgctt 120  
ttaggatgtg aaaaatgctg tatcacttcc atttggcctc catagattca gatgagaaat 180  
ctactgtcat ttgaattggt ttatccttat aggtaagggt aatttctctg tcattgcttt 240  
aaaattgggtg tgggtttttt tgtttttttt attttttctg agattcgttc tcctgagtag 300  
ttgggatgac aagcatttgc caaccatggc ctggctaatt ttttttgat tttttgtaga 360  
gacagggttt caccatgttg ctgagggtgg tctcaaactt ctgagctcaa gcaatcagcc 420  
tcccaaggta tttggtatta caggagcaac ctatctcgcc cagcctaaaa ttgttttcgt 480  
tagtgt 486

<210> 84  
<211> 548  
<212> DNA  
<213> Homo sapien

<400> 84  
tcagcaagct ataaatcctg tttaaactctg tcacgcgcac gtatgattat ctgaaatcgg 60  
cttacgtggt cgggcgagta cttttttctt cttttaataca gatagctgat taacattgaa 120  
acacaaagtc tcattctctc tctaattata ctattttatc tttccttgtc cttttgttat 180  
tttccatata cattactact cttttaataa gaaagttttg cattcttaat ccaactgatat 240  
gttttttccc aatctagctt agtcccatta atttttaact cttttcctgt ttagagtaat 300  
cagttactgg atctacctac tgtctaagtc atctgtcata ttaattgtct ataacttttg 360  
ataaatcagt ataaagctga acattaacat aaatatgtcg ggagacaaac gtttacgttt 420  
ctgaagactc tgcattgtta ataataaaa ccgatatcat ctctttttga tatatactca 480  
gtagtggcat tgtagatta tagtggtagt tatattttta gttttctgag gccctcttta 540  
ctagtttt 548

<210> 85  
<211> 718  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (348)..(639)  
<223> a, c, g or t



<400> 85  
ttttaatctg gcccaagaat gacacaattt cacttgcaact atattaatgg tggccaaggc 60  
agtcattgggc cgcgccagat ccaagggaaa gggagtcag gagtctgtgg tggcttgaac 120  
ctctcacacc aactgattag tatccaggcc cacgcttttag cacttatcag ttaagggtgtt 180  
ctggcatccc cattttcttg gaggagacat tctcctatta ttgagagtta ggtagctcag 240  
acaatcttct acggagatta tgacatgatt ggttgaacac aactagtctc cagtttcttt 300  
ggagttttct ttttaggtat ataatgcaac gtaagtgtct cagggttnnn nnnnnnnnnn 360  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn agaaaagaaa agaaaagaaa 660  
aaaatgaata atcataacc tacatacaaa agctaaaact ggcccatgcc tataatct 718

<210> 86  
<211> 236  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (74)..(74)  
<223> a, c, g or t

<400> 86  
gcaggtaacct ttctctttgt tatgtctcct tcctcgggtg taccagtaat gtgtagtttt 60  
ataccatttg ttangttttc tcatagttct tgggctatgt gggccctatt ttttctcaat 120  
ctttttttcc ctttgctttt tagtttttaa agtttctgtt gtcatatcct caaactcgga 180  
gatccttttt tcatTTTTTT ttacctttt attttggaat tgggggtaaa ccgaac 236

<210> 87  
<211> 587  
<212> DNA  
<213> Homo sapien

<400> 87  
tttagaggta taccaacgat attgggggca acccagggtc cccaccgac gttaccacgc 60  
gccagtgat ttgaatcgct cattatggcc ataatggccc tctagagcaa gctcgacgcc 120

46

ccacagtgtg atgaattctc aagtattcgt cttagcgtgt ccccgcccga gttactatcc 180  
caaacgctgg ctctgctcct tgtgtatatg ggtgacctcg acaaagtcaa tttcaaacta 240  
tctcaagcac tcagtttcct caatttgtaa aatgaggata aataatgtga cctcacagct 300  
cacaggctgt agtgaggatt ccacaagata ctgcatccaa atcactagtg tcttgctcac 360  
ctcatagtag gcatctcgcc aaaatacaca acctatattt ttatcaccac cggactcata 420  
cacatattgg attgtgagaa acattaaacg aagatagcat ccaataaaca ttgggtatct 480  
ttctaacacc ttttcccccc ttccccttac ttcaaatac cctaaccatg atctccatct 540  
taccacccac agagaagacc ctcaaagact ggcctttccc aaactgt 587

<210> 88  
<211> 412  
<212> DNA  
<213> Homo sapien

<400> 88  
aaaggaaaa agagaaagta ccagtggagac gagctccatc gctaatacag gcgcagtgtg 60  
ctggaatcgg ctttcgagcg gcgcggggca ggtgcactcc agccagggct aacataagtg 120  
tagtgagact ctgtctcgaa aattaattaa ttaattgaat caatgtttaa atgtgataga 180  
ataaatgaat aaaccttatt aagatctttt acaacattcg tataagattt tcttagtctt 240  
aacattttta ttgatgactt gatgtgaagc tatattttta aaattacaag attgccagtt 300  
attgcaattg cacacatact gagacaatgt gcttatggat agataaaaaat acaatgaata 360  
aaagggacat tggccaata ccaataaaac aaaattta atcactaaat gt 412

<210> 89  
<211> 843  
<212> DNA  
<213> Homo sapien

<400> 89  
gcatgttagc caactgtacg agtcggacca tataacgccg ccatgtgctg gaatcggatt 60  
tcgagcggcg cccgggcagg tgcttattct cacagattat ttctttaaat agtcttaatg 120  
cagacaaatt gatgattagt aagcccaa atagtataat acagaggtct ttgttatatt 180  
tagcttttta ttcttccatg tgtccccgta ccatgccttc agaacttcaa ctcatatata 240  
atgtctatgt acagaagtaa aacaattatc agcgatgcaa actgaaaaag tctgtcaatc 300  
at ttgggtat gtttatgtta ttgcatattt gttatggata cccttgataa gcaaataatt 360  
gtgggtgctt at ttgttata atgcaaatat tagatatgta aatctagaaa tctttattta 420  
atagctatgt gataagaaat ctcaagttcag ctgtagacaa gaaaacaaat ctcaatacct 480

```

attatattac cataaagcta tttaatctac tatctttcat ctatttcaat gactagctat      540
aatatgttat agtcaaactt ggttacaagt agcctttttt acacattaaa tttgtgtcat      600
ctttcatctg acatattcaa cttataacgt gttatctttt gctgcaatcc taggcattcg      660
agatgcttta gttagaaaaa aggagatgca taacagtgtc taggaatact ggctcaatag      720
tgtacctggg gcggaaccac ctaagcgaac tctggagata tcctaaactg gggcgacaaa      780
cttaattaag ggccgatcgc ctaatgagtc ctaccttctt ggcccggtac actagtggga      840
gcg                                                                    843

```

```

<210> 90
<211> 454
<212> DNA
<213> Homo sapien

```

```

<400> 90
acttaacagg caataactat tagcagaaat agtgcataac taaaggttcc aacgattatt      60
ataaatcaaa aactaagcaa tctagaattt ctaaggcaat tatttaaaat atatagatac      120
taatttacac attaaatggt cccttagtta tttgtgaagt catttacaat gtataaattc      180
acaacttata taaaaacttt aaatgtataa ttctagcttg ccacttaaat aataattggg      240
atgaatgttt tataatacaa atactcgtcc ctttaatttat cagatttttt aaatactctt      300
tttatgctgc tagaagggtt tgtctttgtt ctattactca agctgtggta ggaagaacac      360
catacaataa ttaggaggcc tggattttca tctagttaat agtaaccgga ttgcagcaca      420
ctgcggccgt atattgaggc aggccgtacc tggc                                                                    454

```

```

<210> 91
<211> 757
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (258)..(697)
<223> a, c, g or t

```

```

<400> 91
ccccatgcag aaattactcc gaaggatgca caacgaccta catgtagaaa taaaccata      60
attcttaggg ataaatcttt gtgaccttgg gtcagccaat tatttcttgg attcagcacc      120
agcgacacaa gtaacccaac agaagacgga taaattgaac ttcaagaaaa ttttgtaaaa      180
attgtgcttc aaaggatacc atcaagagag tgagaagaca acctacagaa tgggagaaaa      240

```

48

tagtcgcaaa tcatatannn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 660  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720  
ccagcacact gcgccgttac actgatgcag tcgccag 757

<210> 92  
<211> 667  
<212> DNA  
<213> Homo sapien

<220>  
<221> misc\_feature  
<222> (73)..(567)  
<223> a, c, g or t

<220>  
<221> misc\_feature  
<222> (72)..(567)  
<223> a, c, g or t

<400> 92  
acgaattcaa tgcaattcct atcaaaatat caatataatt tctttttgag atggagtctc 60  
actttgtcat gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540  
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600  
aatggaccta ttaactaact gttgcacgca aatttatttc acgatgacat gcacttgtac 660

tactgag

667

<210> 93  
<211> 581  
<212> DNA  
<213> Homo sapien

<400> 93  
cgagcggcgc cggggcagga ctaaacagtc gttgtaacaa aactgtgggt acaaagtcca 60  
gagactctta aggcaaacaa tctgggaagt agaaatgtca gatttactta ccaaaataaa 120  
aatcatcagt gacacgatca gagaaaaaaa aaaaagaaac gaaaaaaga aaaaaagaag 180  
accctcctcg ggcgagcgaa caagctccta atgccccgag acatctctca ccaccacacg 240  
gcggcagcgc acggtacta gatgggactc accagaacgt ctgcgagaca gagcgttgga 300  
gagtatacgt gtgatctaag agtggttaatc ggggtagaaa cagtgcgtca accaggcctc 360  
cagatagccc cggaacgaac agtctaccag acagccaaaa cgaaggaagg cgagagggga 420  
ggcagcgagc ggcagacgag agagagaaga agaagagaga gagaagagag aaggaggggac 480  
gaagagagcg gcgaagggac gcgcaaaaga cgggagggaa gggcagcgaa gagaacagca 540  
ggagaaggag ggagaagagg ggggggaagcg acgagggagg g 581

<210> 94  
<211> 619  
<212> DNA  
<213> Homo sapien

<400> 94  
gttattgtaa gagataatct aatgccccag atggcctttc catagttaca gctttttattg 60  
ttctcaaaat gtgtgtgttc tttatttgtg tttatgggaa gaataaagat tttcttgttt 120  
gccagggggc atgtaaggta accctggcag gttagtgtgt gtaagaatac acgggaataa 180  
ttttctttgt tactttctag gacaactcat aattttattc tggaagtcac cttagttcct 240  
gtggttttcc ctcgacaatt tgaaatttct ggctccagaa ctccggattt taagcttgat 300  
agttatttct catgatattc ctcataagct aggtaatatg agattttaag ccatggatta 360  
aagaataaaa aatgttaagg tttggaagta gtcttatctt tttaactttg atagttcata 420  
ttttatatct cagccttgga agctgcaata ggatggtgta tgttctcaaa gcgacacttc 480  
gcaaatttat ttcgtacttg tacacaacag gtgacttgta caattctgta accaaattcc 540  
catggattgt gcagaagaat caattcacat tttcttatta acaaagtctc agagtttaaa 600  
aaatacatat ttttacagt 619

<210> 95  
 <211> 544  
 <212> DNA  
 <213> Homo sapien

<400> 95  
 actaccagag ataggcactt taataaggat aaaaatggca aactggataa tggttaatgat 60  
 cctaaatttg aaaatatcta ataaaaactt caacatacat aaagcaaaaa ctgacaaggc 120  
 taaaaggaga aataaggaaa tccataatca taatgggaga ttttacacat ctctttcaga 180  
 aactgatata tgcagacaaa aattagttag aatacagaat atgttaacac aattaaacaa 240  
 aatggacaca cctcgagcag tatatctagt aaatgcactg ctacatgtcc tttaaaata 300  
 cgaatagaac atttgcaaaa tgaccagctg ctaggccatc actgccaaat gctcaatata 360  
 ttctaaatga atctgtaata cagcactatg ttccacatac agaaagatat gttataaaat 420  
 ccccatcgtc ttggaaaata atatacttcc tcaaaaatgc atgggtccaa gaagaaatcg 480  
 aagtggatat atgacaatat cctgcaagat aatgaaaata ctacacatcc taactgatcg 540  
 aggt 544

<210> 96  
 <211> 588  
 <212> DNA  
 <213> Homo sapien

<400> 96  
 gaagaaacat aagcgactgg cttaaatcat gtcacggcgc agtgtgatga tattgcagaa 60  
 tcggcttaca gcgataacag gtatgagaca ccgccattat caagctgggg aatttttttg 120  
 ctaaagttag tgggtgagaa actgacttta gtatagttta tcttgcaattt ctttattagg 180  
 agcgatgttg aacatcttcg caacgtgttt aaagaccagc ttgtgtgtat ctccttttgt 240  
 gtaaatgtgc tgttcttgtg catgtttgccc tattttgtct gctaggattt cttgggtctt 300  
 gtttctgtga agttttataa gcttctttac tcagttaggg actactgagt tttgttagtt 360  
 tgcaatgtag gctgcaaatg ttttcttcta gggttgttgt ttgcctgttt gactttgctt 420  
 atggtttttg gcataccaaa gtttagaaat ttttataaat tatatttata aattttgttc 480  
 ttgttagtgc agtttgaatt ttttagtaaca gaaaacgctg tttctgatat ctagattaca 540  
 gaggtagtta tcttatgttt tctttagta tgctggggcg gacgatgc 588

<210> 97  
 <211> 514  
 <212> DNA  
 <213> Homo sapien

<220>  
<221> misc\_feature  
<222> (102)..(132)  
<223> a, c, g or t

<400> 97  
ctttaacatg aataacaaat gcttttgaca gactcaatac catactttat acagctgaag 60  
aaagagtatc taagcataag tatataatga tagaaacttc cnnnnnnnnnn nnnnnnnnnn 120  
nnnnnnnnnn nnggaaaaaa atcagagcaa agtaatcaag agctatagga cagctatgaa 180  
aggtataaca tacatgggaa tatcaggagg agaagcaaag agataacagg aatagaaata 240  
atatttgaac gtaaagtact acctgagatt tcctgaatta acagtgagac actggaacga 300  
acaggatact aaacaaccta agagaatact aaagctgaat aaaatgcaa gaaaagaaga 360  
aaactacata taagcatatt atattcaatt gccgaaatca agataagaaa aatgttgaca 420  
gaatccagag aagaaaaaaa cttagaagaaa agaaggaagt tagatttttg gttctttgaa 480  
acagcaggaa agaaaggggg ttttgggggg aaaa 514

<210> 98  
<211> 1300  
<212> DNA  
<213> Homo sapien

<400> 98  
acaggatatg taaaataact gggattatct tatgagaagt ttgcactgtg agttacatca 60  
tgacattaat ctgaacagac cttttttgat taatacacat ggaaaataca ctataaattt 120  
tagaatgttt gtgttttgac tgtggaaaat ttacatattt tatgtatctt tagataactt 180  
agctcttaaa tttagtgtt caataatttt gccattttgt aaaaacagat atcttccatg 240  
catctacca actgactgaa acttatttat tcctatatga ttactttaat tacaacata 300  
atcatgaaaa gcttttgaat gttcagatgc acagtctact tcctggatac tttgctttta 360  
ccgcacgtcc ctttgtttct cagtgtttat ttggccttta ttataaagca aaaatatgga 420  
acattttttac tcctgtggag acataggctt ttacctcgtt aacctcttat ttaaaacttt 480  
tatcacctat tctgacaatt ttcttaaacg acaaattata ttcaactatt taattcttag 540  
aaaaatgcc aacacattttt aatccctaaa ctacaactgc attgtcttga atccttttgt 600  
gaaatatgtt ctttttttag cctttaagat catccagtag gcttcatgga aataaaagac 660  
taaaaggatt attacttgac catctttaat cagatgttgc ttgagagtcc tagaaaaaaa 720  
tatgccatta ttttaggagg ttgaagaact gaataatttt taaagtcata tgggcagtca 780  
gagaagaaaa tcttttagat gtctctaccg aataatatta accacctaga aaaaaagcga 840

ccacactgac aattttcttt aaggactgag aggagattat ttcaattacc atgttgcctt 900  
 gtccttgtca actttttcta ggatgtcaga gctaagcaaa taccottaag atacacttaa 960  
 aggagaggct ggttgaaata agagaatcta aaagttgtct gcaggcttaa gaatgtctag 1020  
 ctttaaaaat attaaaccgt gaaaaagttc tgccaactat gtatatgttc ctccagctat 1080  
 tgttctactga cttagtttat aactctgcat gggactgcgt ttcttctgtt cacaaggaat 1140  
 aatgtccaat aagttgtaaa gaacaatttg gaccatattt atgtgcaaat atttcattga 1200  
 tatcatcact taatgattct aatatttccc atgtgtcagc ttgtgcactt tttgtgtaga 1260  
 attctgtgac cagaagataa aaatctacta gcagatatatt 1300

<210> 99  
 <211> 340  
 <212> DNA  
 <213> Homo sapien

<400> 99  
 acaacaaacc cttgtgacat gcaatctatc taaataacaa acttgccaaa aaaaaagggtg 60  
 gagatgaaat actcaatctc tctctctctg ctttttccaa gtattttttc tttcaggggtg 120  
 ctaaataaaa tatgttaaca tataatgcct tatttacaga aataaaagtt agttattgaa 180  
 acaatttgta atttagttct tgaagtaaga aatacaaaaa ggtgataatt agggcagttg 240  
 aaaatcagca ttatggggat agatgtgttt tccttttggt tccaagagct gggcagacca 300  
 tgctgaagag attcagttct ctgaatttgc tccaagtgg 340

<210> 100  
 <211> 888  
 <212> DNA  
 <213> Homo sapien

<400> 100  
 cctttcttca taaatttgaa tagcctccat gaaattgagt aatctcataa aaatttaaaa 60  
 ctgggttttta agaacatatt ctgaccacta aataaaagtg cttctgtagt tacatgagtt 120  
 aatttgggag ggagagcaga acgagtttca aagtagttgg cctagcatag tcaattagat 180  
 aaagcagtgct attgtaattc aaatcaattt tcattttgta attattgcta tcagtaaaca 240  
 tttttttttc aacatgcagt gcagttaatt ggtaagggtc agcattaatt tactcaacat 300  
 atatgctgct tttggagtta aaaatctttt ctcaaataaa aaactcctga actgtctgtc 360  
 tttattctca gtattgctaa tgactaagca gatggctgct gttgaaactt cttttcctcc 420  
 actacctgtt tctgtttata ttctgatgaa tgcagatata gttctagtgg cattctcagc 480



53

tgatacgggtg ttgacttctt ggaagtttgg caaaacctca gggaacaatt tttcattgcc 540  
 agtttttaaag ctgttttagga cctttatcta aaaaacttaa agtggttacgg aatttcttcc 600  
 acatttttaga gatgccttta atatcctaca tagtcatcat atgtatacag ctcttgccat 660  
 acttagcaga ttgggtagtc taatctccag taaagttgga attgtcatct attaaaatac 720  
 tggaatcaca aagcaaatgt ctaatttaac gcaccttaat ttttaatgat gtgaaatcat 780  
 ggataaatatt gtgattttctg agcatctaatt tatttcattt atgtacctgc ccaagaccga 840  
 attgcagcac actgcgccgt attcagcgag gtgagctcga tcaactgga 888

<210> 101  
 <211> 937  
 <212> DNA  
 <213> Homo sapien

<400> 101  
 tctaaagtaa atattacctt ctcatTTTTg tgaaccttca ttattttttac ttggttaagtt 60  
 aatgcaactg cgaaatcagt actcctttct tcataaattt gaatagcttc catgaaattg 120  
 agtaatctca taaaaattta aaactggttt ttaagaacat attctgacca ctaataaaaa 180  
 gtgcttctgt agtcacatga gtttaatttg gagggagaga gaacagttta aagtgttgcc 240  
 tagatagtaa ttagataaga gtgattgtaa ttaaataatt ttatttttaa ttttgcttag 300  
 taaatttttt tttcaacatg cagtgaatta attggttaagg tccagattaa tttactaaca 360  
 tattgctgct tttggagta aaaatctttt ctcaaatcaa aaactcctga actgtctgtc 420  
 tttattctca gtattgctaa tgactaagca gatggctgct gttgaaaact cttttcctcc 480  
 actacctgtt tctgtttata ttctgatgaa tgcagataag ttttgtggca ttctcagctg 540  
 atagtgttga cttcttgga gtttggcaaa cctcagggaa aatttctatt gcaagtttta 600  
 aagctgttta ggactttatc taaaaactta agtggttacgg gatttcttcc acattttaga 660  
 gatgccttta tctctacat agtcatacata tgtatacagc tcttgccata cttagcagat 720  
 tgggtagctt aatctccagt aaagttggaa ttgtcatcta ttaaaatact ggaatcacia 780  
 agcaaatgtc taatttaacg caccttaatt tttaatgatg tgaaatcatg gataatattg 840  
 tgatttctga gcatctaatt atttcattta tgtacctgcc caagaccgaa ttgcagcaca 900  
 ctgcgccgta ttcagcgagg tgagctcgat cactgga 937

<210> 102  
 <211> 542  
 <212> DNA  
 <213> Homo sapien

<400> 102  
aagaaaaaa ggaccaggta gtacgagctc gatcattgat acggcgcagt gtgctogaat 60  
tcgtgctttc gagcggcgcc cgggcaggta ctatcatatc ctataaaaac acaagttttt 120  
aacataaaatt gtcataatattt gccaatcttt cccactctg gagctgtctt tttttttttt 180  
ttttttaaag tcaagcacia tgttttaaat ctccaaagtt gttgaaagga aaattcagtt 240  
tattgtttta attttgatta aaacttttga ttagttgaca gcgcaacttt taaattatatt 300  
tttacttcat ctccctaaag cattgagttt gaattttctc ttggtgcttt tgttgagatt 360  
ttttacaagg aatgggtgtc aattaagaaa tatgattact gcgttagtaa cataaaaaat 420  
acttatctga aatactgagg gcacaattag tgctatggaa ctggatattt acatacctac 480  
tctgagaggc atgggaaatc tctgtgtctt tcttctgtac ctgcgcgca ccacgctaag 540  
cc 542

<210> 103  
<211> 793  
<212> DNA  
<213> Homo sapien

<400> 103  
aagaaaaaa ggaccaggta gtacgagctc gatcattgat acggcgcagt gtgctogaat 60  
tcgtgctttc gagcggcgcc cgggcaggta ctatcatatc ctataaaaac acaagttttt 120  
aacataaaatt gtcataatattt gccaatcttt cccactctg gagctgtctt tttttttttt 180  
ttttttaaag tcaagcacia tgttttaaat ctccaaagtt gttgaaagga aaattcagtt 240  
tattgtttta attttgatta aaacttttga ttagttgaca gcgcaacttt taaattatatt 300  
tttacttcat ctccctaaag cattgagttt gaattttctc ttggtgcttt tgttgagatt 360  
ttttacaagg aatgggtgtc aattaagaaa tatgattact gcgttagtaa cataaaaaat 420  
acttatctga aatactgagg gcacaattag tgctatggaa ctggatattt acatacctac 480  
tctgagaggc atgggaaatc tctgtgtctt tcttctgtac tggagggttt ccatctgcag 540  
tcatttcaat catacaaaca actccagggc tgccaacagg cgttgtgtag ctttttgcag 600  
ccttttcatt tgtttttagat aacctgaact tccatttggg cagggtcttct aaaaatctac 660  
aaataaattg tcttttccaa atcagaatgt ttccaccaat attgctaagc atacaggcag 720  
atctgtggat gagctcctta ggatgagcac ctattgccaa tgaaatatct ccaggatatt 780  
gtcattttg cat 793

<210> 104  
<211> 829

<212> DNA  
<213> Homo sapien

<400> 104  
 tccagtggtc gagctccatc actgatacgg ccgcagtgtg ctgcaattcg gcttacttgg 60  
 ttatcttaaa aacactgtga aactagtaga atgttggtt aactgtagat attcttattt 120  
 tttccaatct tagggataat tcttattcct ggcagtgtgt atttactcag agagcattat 180  
 ccagctgggg gtttcaactg taaagctcag attggtctac tgataccctc ttaagtagta 240  
 gtaatttaat acacctaaat tccatgttct attgatgtgc acacgcaact aaaatctatg 300  
 atcagggtttt agtattctgt tagttattca actggatggg ctctgtccgc ataaaatgtaa 360  
 tagatcatga atcagccaac gaattttgag aaggagaaaa taacaatata tgtgtattat 420  
 gtgggttatat aacaatataa caatacaacc tatactatgc ttctgctgat ggaatcctct 480  
 ctactcatg catgtattcc tgccctctcg ctgttatcaa gcctatgtac tgacaagcac 540  
 tggatcatte aactatctga ttagaggtct aagggtagtt attccatgag accatataag 600  
 aagggccctc agagaaggag acaatgtaaa caaaaatcat caagggttcc tttctttcaa 660  
 gaatcaacac gtgccagtct cttgcctact attgatcatt atttcaacaa ttgatagtat 720  
 tcttatcatt cgcccactat ttgatggggg tgtttgtttt tttctgtaaa ttatgtttga 780  
 ggtacaagac gaagtgcagc aactgcgcc gctataagtg atggagtgt 829

<210> 105  
<211> 745  
<212> DNA  
<213> Homo sapien

<400> 105  
 accaagaggt aagtttttag taactaaaag ggagaagtaa cttattatta gaaagttoca 60  
 atatatagaa tggggtttca tcagaaaata atggctctta gcagccattg tcttctgcag 120  
 aagagtgtca ctacgggaaa tatattagat agagacgcac tcgactcaca ctccctggat 180  
 gtatagcgct aataactcgt tataaagtc attaataata ctaagcatct agatacaatc 240  
 tttgataaag catcccaaag tccttcaact ctatataaat cttattccat aagaacaatt 300  
 tctatttata atcaaacaca agatgctgca tgcacattta agcattgaga gttacagcag 360  
 taaataagaa agtggaggct cccttgaagg ttacacttat attgggagga gtgggtggat 420  
 acatgggtta caataagcaa tataaaata ggaagaaata gaaagtagga gaaagtcct 480  
 tccccattga cttgtgtgaa accccatgt cactctaggt ataataattt taaatgtgac 540  
 aggtgctgac ctctgtctgc gaggatggct agcatcacct tcttgtctcc ctggactact 600

56

gtttgaatgt ctacattgtc tctcaccact tgcccatgga acccaatttc tatggatttg 660  
tcctgctgat ctctggcaga tccttgggtc ctgagaacct ctggattttc agtctaagtg 720  
ccacttgagt tcctaattctg acagt 745

<210> 106  
<211> 698  
<212> DNA  
<213> Homo sapien

<400> 106  
aagtggtcga gtcgcatcac ttatacggcg ccatgtgctg gaatcggtt gggcaggtac 60  
cagttaatta tcacaacaaa tttcacactt tgcaactatt ggtttatcaa agtaattcag 120  
tgatggctga catggctgcc caaatgactg tgtaatgaat acttcattaa aaaaatatgt 180  
cgctgtcgta gcaatgatga ccttttagat gaggcagtag ggtgatgctc agttgttgtt 240  
ctcataattc tctactaaca atgatgacat gatttatgat tatttacttc aggttgctta 300  
gttaccaagt ctctcttgta taattctttt tgcttgcttt cttctaagag tttttttttt 360  
tttttaaagg aaatagccac acggtttttc tttttttctc tttaaaaata gccatctcca 420  
cactttttgt gcttgaacag aagaaattct ctctccttag ggaagaaaag gggagattgt 480  
agggtgattg aattgggcaa taatgtgtct agagagattt cgtgggtaag gaaaagcttg 540  
attgtgacag acttggtgtt tggaagagag acaaggcaac ctcccacaat gggaattaac 600  
cagcaagagg accttgactg gaaatatcct aatgtcgcta ggaaacaaag ctagagtcac 660  
tttgtacca gtatgagatg taagccttct attggagc 698

<210> 107  
<211> 849  
<212> DNA  
<213> Homo sapien

<400> 107  
ggtgtttcta cagaggtttc ccacgcgatt aaattgggta ccccggggt ttcccccgcc 60  
cgaccttggt aaccgcccgc cctggattgt atacgactcc ctattgggcg cattggccct 120  
cttagaacca tgctcgcccg gccccccctt gtggtggaat ttcccgattc tggcttttgg 180  
cgcccccccc gggcaggtac ccccccttg gccacagagc aacaccttga ctcaaaaaaa 240  
aaaaaaaaaa aaccaaaaaa aaaaattttt ggaaatttgg gcttttgggc cccacttact 300  
ccatgggctt tccagtgtca tagatactg tcatgaaatt tcttttacia ttggcttctc 360  
cagaagggag atacgccttt tggactctag ttgcacaggg ttgctccacg tctctggaga 420  
ctttcgtcag ggcccacatt gtggaagaca gaattgctcc tgtggcgcca cgagtataaa 480

tatataactaa cgtgggctaataat ggatcagttg gatggacgct ttataaaaac aaaacaggtg 540  
tttgggtccca tcaatttttcc aattgtatcc ccacccaacc aacaagattt gggcggagaa 600  
aaagtccttt tgtacgtcaa acgcgtgttg ggcacccggg ttggtctotaa gtggtctttg 660  
ccggcccgtg tttggggccc cttcggggccg ggcgaccaac cgtttaagcc gggatttcca 720  
acgccccatt gggcggcccg gtacactaag tgggtctccc gagttctggt ccccacagtt 780  
gtgcgggtac cccttgggccc caaacgctgt cgctgtgtgt gaaattgttt cccgtcgatc 840  
acaaagcag 849

<210> 108  
<211> 605  
<212> DNA  
<213> Homo sapien

<400> 108  
acattgcagc attttaagaa cagaatatgt agatgtatgt acaacttctc tcttccccc 60  
gctaaggagc gttaattata gcatgtaacc taagttattt tccaatttta gaaaattagt 120  
ttagtatatt cattttctgc acatttgaaa tgtaattgga tatcttttat tgtcatttta 180  
atagtataac tatttctggt agcctatatt tttatggctt aaatagctct ataaacttta 240  
gggttttttc caagtctgtg ggaggaaaat acattggctt cctaaaccat ttagatccag 300  
accaaataagaa tataattgca tcttttaaaa aatattaagt ttgtaaaaat cattgggaga 360  
actggtagca ttttaaagag gccctatag taagctgtga tggcagcaaa taacttttct 420  
gggttggggg acgagcgtct ttctgtcaa acggggcaaa tagaacgaca tactactttc 480  
tggcaactaa tttatttcct cttcactctc ttctattgaa gatctaggaa agaaatataa 540  
tgtatacctt ataactctgt gaaatttttt tctgtgtgta gactgcgatt aaaaagtgat 600  
accca 605

<210> 109  
<211> 959  
<212> DNA  
<213> Homo sapien

<400> 109  
agagtatccc tcttatttcc ggtgtcctta gtagtcagaa gtagcaaatt aggatggaga 60  
atgatcataa tagcaaataat gtgttacatg ctctctatgt gccagaaact gctaagtgtt 120  
ttatatgtgt ggtttcctta gaggcaagta tattaatagc agccagaagt gacacctgat 180  
ggcaagggga gaaaacactg ttaagaccag aacagtctga accttggtat gaggcaagca 240

58

gatctaccta ctcagcccaa tgtctgtccc tgtattccag cactgcatac atgaggacat 300  
taataatcaa gaaaaggtta aattacattt ctcccttggtg gtttaaatgta tagtacattg 360  
cagcatttta agaacagaat atgtagatgt atgtacaact tctctcttcc ccagagtaag 420  
gagcgtaaat tatagcatgt aacctaagtt attttccaat tttagaaaat tagtttagta 480  
tattcatttt ctgcacattt gaaatgtaat tggatatctt ttattgtcat tttaatagta 540  
taactatttc tggtagccta ttttttatg gcttaaatac ctctataaac tttagggttt 600  
tttccaagtc tgtgggagga aaatacattg gcttcctaaa ccatttagat ccagaccaa 660  
tgaatataat tgcattcttt aaaaaatatt aagtttgtaa aaatcattgg gagaactgg 720  
agcattttta agaggccct atagtaagct gtgatggcag caaataactt ttctgggttg 780  
ggggacgagc gtctttctg tcaaacgggg caaatagaac gacatactac tttctggcaa 840  
ctaatttatt tcctcttcat actcttctat tgaagatcta ggaaagaaat ataattgata 900  
ccttataact ctgtgaaatt ttttctgtg gtttagactgc gattaaaaag tgataccca 959

<210> 110  
<211> 788  
<212> DNA  
<213> Homo sapien

<400> 110  
ggcgccatgt ctgcaattcg gtgggcaggc cttttattta ttttgtaag agtaagtgc 60  
ttttgacaaa ctttcagttt gttaccaagg aaggaaagaa atagaggatc tcttctgatg 120  
gttacatcat tactcctttt ggtaacaaa accacttatt actttccctt tgggtagcaa 180  
tcttgaattg aagttcttca cagatgacct aaaatggcaa tctgtcctct aagaacttgc 240  
attttgagag atgacaaatg tcccaggta acttaacatc aaataggta tccactgggtg 300  
caggccactt tacaatatgt aaagggtttt tacaacacct tgttcaagtc tcatgataac 360  
tctgtgaaat aagttgggccc aatattgtca ttatcctcat tttatgcgag gacaaggaca 420  
gtgagtcttt gtgatgacag gggcctctct agcataaagc agatagaagt gtgagcactc 480  
cggcataatta tttatttttg ggcttggctt tttatagcaa attatcacta gaaaggcatc 540  
catctttaga atttcttttc aaacaagata tttccaaagg tgttgtaac ttagcccatc 600  
tttaaagtct ctctgtttt gactggatct gggcatggct tgtacctcg ccaagcagaa 660  
ttctgcagat atcatcacac tggcggcgct cgacatgctt taaaggccaa tccgccttta 720  
ttgagtgtat tcaattactt ggcggttta aagtctattg gaacctggtt ccattctcac 780  
ttttaaaa 788

<210> 111  
 <211> 335  
 <212> DNA  
 <213> Homo sapien

<400> 111  
 tccaagactt ttacttcaact gtctttcaat tgcttgcgta tggatatccct gaggactgac 60  
 gatactcaact gacatgaggg tagggctggg gaggagcgct cactatataa gtcactgagt 120  
 gaatgagtaa atcataatTT tacctcccca ttttctccta ctcttctcta atcacctatt 180  
 ctttgtcatc cccaactacc ttcatgattg gaaagataac gcgagttggt gagaaaaaaa 240  
 cactaggttt ggtttcagtg ccccagaagt cctaactcta ttactgtaag ccgaagtcag 300  
 cacactgcgc cgtataacgt aggcggctcg ttcct 335

<210> 112  
 <211> 1101  
 <212> DNA  
 <213> Homo sapien

<400> 112  
 tcgcgctcgc ccactatacg gcgcagtgtg tggatcgggt gggcaggtagc gtggtagagt 60  
 atgctacgag taaaaaattg ggagattcag actcaaattc tcttgagatt agaccaaagt 120  
 atatttatca aatgcttagt tgggcataaa aatactccaa ttacagaact tgcattattat 180  
 tatccactct ataatagtag agagtcataa cctatgattt gtccccttgg aagagtctat 240  
 tcattatact ctaaaagtta aaccactggc ccttatatgt tccatatatt atcagattca 300  
 tatatagaga taattaacct attccttatg gatgtaaaat aggcgtttca aaattaacat 360  
 agttcaacaa ttgaactctg gagttctctt tccatagttt attctttctc tgatctttcc 420  
 cgtatttgta agagtatttc tttatactaa caccaaaaag cacacaatca tcgggacaat 480  
 atctatatac tttagctctt ataccacagc ctgtatcctt caggcaaatac ttgccatcct 540  
 aaacttgtcc ttaaatacct cagtccccct tttgtgcta ctctagtcca ggccaccata 600  
 caagtctcat ctttaactac tgcaatagtc tctgttgaaa gattctcctt atccagtact 660  
 ctacctttgt ttaccatttt caagtgagct tctaccaag gagtagtacg gaggtocata 720  
 atactctgcg tggaaaaacc tgtcctcagc acaactactt aaactcacat taccataagt 780  
 attaatgata actaacagcc tacctttgta atattccaga gtaaaactatc acttttaaaa 840  
 agatataccc ttcaactggg gaattaccaa gcatgtggaa gagttttttc ataacaactc 900  
 ttttcatgac ggattttgac acagcccttg gaatttcttt ttttaatgat tgaaactaac 960  
 cctgtttcac tcaccctctt tttcacaaac aggtaacatc ttctttctca gagtaatata 1020

gaggataacg atgacaacac atgaaataaa ttaaaatggt atgagtgcac ctataggtag 1080  
 acaaatagaa gaactagatg t 1101

<210> 113  
 <211> 1181  
 <212> DNA  
 <213> Homo sapien

<400> 113  
 tcgcgctcgc ccactatacg gcgcagtggt tggatcgggt gggcaggtac gtggtagagt 60  
 atgctacgag taaaaaattg ggagattcag actcaaattc tcttgagatt agaccaaagt 120  
 atatttatca aatgcttagt tgggcataaa aataactcaa ttacagaact tgcattat 180  
 tatccactct ataatagtag agagtcataa cctatgattt gtccccttgg aagagtctat 240  
 tcattatact ctaaagttaa aaccactggt ccttatatgt tccatatatt atcagattca 300  
 tatatagaga taattaacct attccttatg gatgtaaaat aggcgtttca aaattaacat 360  
 agttcaacaa ttgaactctg gagttctctt tccatagttt attctttctc tgatctttcc 420  
 cgtatttgta agagtatttc tttatactaa caccaaaaag cacacaatca tcgggacaat 480  
 atctatatac tttagctctt ataccacagc ctgtatcctt caggcaaact ttgccatcct 540  
 aaacttgctc ttaaatccct cagtccccct tttgtgcta ctctagtcca ggccaccata 600  
 caagtctcat ctttaactac tgcaatagtc tctgttgaaa gattctcctt atccagtact 660  
 ctacctttgt ttaccattt caagttagct tcctaccaag gagtagtac gaggtccata 720  
 atactctgcy tggaaaaacc tgcctcagc acaactactt aaactcatat taccataagt 780  
 attaatgata actaacagcc tacctttgta atattccaga gtaaaactat actttaaaaa 840  
 agatataccc ttcactgcgg gaattaccaa gcatgtggaa gagttttttc ataacaactc 900  
 ttttcatgac ggattttgcc acagcccttg gaatttcttt ttttaatgat tgaaactaac 960  
 cctgtttcac tcaccctctt tttcacaac aggtaacatc ttctttctca gagtaataca 1020  
 gaggataacg atgacaacac atgaaataaa ttaaaatggt atgagtgcac ctataggtag 1080  
 acaaatagaa gaactagatg tacatctact catttgatga tgactcctca agctttggcc 1140  
 atgcgtagac tagtcagttt ccagttgtgt gactagagca g 1181

<210> 114  
 <211> 552  
 <212> DNA  
 <213> Homo sapien

<400> 114



61

```

accacaaaac attttttgta aagaaagata tgacaaagaa aaaataattg ttgttattta      60
gagatcgtgc gtaatctatc ttaataaaac agatgaagga tatacacaca tgaagaaagg      120
gtgtgatggt taattttatt tggtaacatc ctgactgggc cccaaatatt cacttaagtt      180
attattcatg tggttgttct ggatgaaggc gtttctgtga tgagattaat catttgtaca      240
tcggtagact ggaataacag cattattttcc ctccctaatt tcgagtggat gcctcatcca      300
attctgttga agttctgtaa tagtaataga atagctaaca tcaatatttt tttctgtct      360
tgcaatgcct ggactgtctt tgaagcccta ggacactggg ttttctctgt gccttttttt      420
tttatttttc tatttttggg gggagaggaa tccttttttt ccaaaacaaa acaaaaagg      480
cttttataat tggggaaaat tccattttgg ggggagtaaa acttaaaaat gtggaagggg      540
gggggggagaa at                                                              552

```

<210> 115  
 <211> 44  
 <212> PRT  
 <213> Homo sapien

<400> 115

```

Met Val Leu Asn Ser Leu Pro Ser Leu Cys Thr Pro His Asn Ser Thr
1              5              10              15

```

```

Cys Ser Trp Leu Leu Thr Pro Asn Pro Cys Ser Ser Leu Trp Lys Gly
                20              25              30

```

```

Phe Leu Leu Val Tyr Val Arg Ile Gly Leu Lys Cys
35              40

```

<210> 116  
 <211> 62  
 <212> PRT  
 <213> Homo sapien

<400> 116

```

Met Glu Thr Phe Phe Phe Ile Lys Ile Phe Trp Leu Thr Glu Tyr Arg
1              5              10              15

```

```

Ser Asn Lys Asn Lys Arg Asn Asn Gly Phe Arg Asn Leu Leu Val
20              25              30

```

```

Val Ala Thr Ile Tyr Ile Thr Lys Arg Glu Ser Gln Ala Asp Leu His
35              40              45

```

62

Val Leu Arg Lys Ala Val Asn Ile Thr Tyr Asp Leu Ile Cys  
 50 55 60

<210> 117  
 <211> 38  
 <212> PRT  
 <213> Homo sapien

<400> 117

Met Tyr Ile Leu Arg Thr Leu Lys Thr Ile Lys Asn Ile Met Ile Thr  
 1 5 10 15

Ala Ala Lys Ser Asn Lys Leu Phe Asp Ile Asn Ile Tyr Pro Val Gly  
 20 25 30

Ile Lys His Ser Ser Tyr  
 35

<210> 118  
 <211> 31  
 <212> PRT  
 <213> Homo sapien

<400> 118

Met Gly Lys Ser Gln Gln Ser Asp Lys Arg Lys Lys Glu Arg Ala Ser  
 1 5 10 15

Asn Trp Lys Thr Gly Ser Ile Asn Thr Ile Val Ala Val Cys Gln  
 20 25 30

<210> 119  
 <211> 65  
 <212> PRT  
 <213> Homo sapien

<400> 119

Ala Ile Arg Gln Glu Lys Glu Ile Lys Gly Ile Gln Thr Gly Lys Glu  
 1 5 10 15

Glu Val Lys Leu Ser Leu Phe Ala Asp Asp Met Ile Leu Tyr Leu Glu  
 20 25 30

Lys Pro Arg Leu His Gln Lys Thr Leu Glu Leu Ile Asn Lys Phe Ser  
 35 40 45

Ile Val Ala Arg Tyr Lys Ile Asn Ile Gln Lys Ser Val Val Phe Leu  
 50 55 60

Tyr  
65

<210> 120  
<211> 66  
<212> PRT  
<213> Homo sapien

<400> 120

Met Ala Ser Ser Leu Thr Leu Thr Ala Gln Cys Ala Gly Ile Gly Leu  
1 5 10 15

Tyr Ile Pro Leu Ser Glu Leu Asn Glu Ser Met Asp Leu Phe Gln Leu  
20 25 30

Phe Leu His Tyr Arg Ala Ser Val Leu Val Ser Cys Tyr Asp Cys Phe  
35 40 45

Gly Leu His Trp Leu Asp Asp Cys Ile Ala Trp Asp Tyr His Lys Asp  
50 55 60

Pro Gly  
65

<210> 121  
<211> 26  
<212> PRT  
<213> Homo sapien

<400> 121

Met Asn Ala Val Phe Tyr Gln Ile Val Gly Ile Asn Trp Leu Ala Ser  
1 5 10 15

Ile His Val Ser Ile His Gln Gln Arg Tyr  
20 25

<210> 122  
<211> 48  
<212> PRT  
<213> Homo sapien

<400> 122

Met Glu Met Asp Ser Ser Leu His Asn Ser Met Thr Tyr Thr Val Ile  
1 5 10 15

Phe Pro Ser Arg His Ile Phe Phe Thr Tyr Phe Arg Leu Asn Ile Leu  
20 25 30

Lys Leu Val Asn Glu Ser Ser Lys Tyr Lys Arg Thr Lys Met Glu Lys  
35 40 45

```
<210> 123
<211> 24
<212> PRT
<213> Homo sapien
```

<400> 123

Met Cys Lys Phe Val Thr Trp Val Asn Tyr Val Ser Cys Gly Phe Gly  
1 5 10 15

Ile Leu Thr Ile Ser Ser Pro Arg  
20

```
<210> 124
<211> 61
<212> PRT
<213> Homo sapien
```

<400> 124

Met Glu Gly Ser Leu Ser Pro Val Val Leu Leu Phe Leu Phe His Ser  
1 5 10 15

Leu His Ile Val Asp Ile Phe Arg Val Ile Gln Leu Leu Lys Asp Met  
20 25 30

Asp Arg Thr Gln Asn Trp Tyr Gln Asp Leu Pro Thr Gly Asn Tyr Leu  
35 40 45

Met Leu Ser Leu Asn Ser Leu Ser Leu Ser Val Ser Arg  
50 55 60

```
<210> 125
<211> 82
<212> PRT
<213> Homo sapien
```

<400> 125

Met Arg Glu Tyr Ser Phe Ser Ala Glu Tyr Phe Ser Arg Pro Leu Cys  
1 5 10 15

Ile Arg Ile Pro Gln Cys Ala Phe Met Glu Val Val Ala Ile Phe Gln  
20 25 30

65

Lys Phe Asp Ser Tyr Tyr Ser Arg Gly Ser Val Asp Gln His Trp Glu  
 35 40 45

Asn Val Asp Ile Ser Thr Cys Lys Gly Ile Pro Leu Leu Lys Asp Phe  
 50 55 60

Ser Glu Ser Cys Ser Tyr Ala Gly Phe Phe Asp Ile Pro Lys Phe Cys  
 65 70 75 80

Gly Lys

<210> 126  
 <211> 52  
 <212> PRT  
 <213> Homo sapien

<400> 126

Met Met Leu Arg Trp Arg Trp Ala Gly Gln Lys Gln Ser Ala Val Ala  
 1 5 10 15

Cys Asn Tyr Cys Val Met Trp Ile Leu Leu Ser Leu Lys Leu Ser Leu  
 20 25 30

Leu Gly Tyr Ile Ile Val Arg Leu Gln Arg Lys Ile Ile Ile Thr Thr  
 35 40 45

Gly Gln Asn Arg  
 50

<210> 127  
 <211> 57  
 <212> PRT  
 <213> Homo sapien

<400> 127

Met Phe Cys Arg Asn Arg Lys Ile His Thr Asn Asn Ser Asn Ile Ser  
 1 5 10 15

Lys Asp Pro Gln Met Ala Lys Met Ile Leu Lys Lys Asn Val Phe Gly  
 20 25 30

Gly Pro Gln Thr Pro Cys Cys Gln Asn Leu Phe Pro Ser Tyr Asn Asn  
 35 40 45

Gln Asn Ser Ile Val Leu Ala Glu Arg  
50 55

<210> 128  
<211> 53  
<212> PRT  
<213> Homo sapien

<400> 128

Met Cys Lys Asn Trp Pro Ser Ile Asn Ile Ile His Trp Ile Asn Ile  
1 5 10 15

Lys Phe Lys Ile Pro Phe Thr Leu Gly Lys Gly Lys Arg Arg Glu Ile  
20 25 30

Tyr Glu Arg Arg Met Leu Gly Val Ser Thr Met Phe Phe Phe Asp  
35 40 45

Phe Phe Met Ser Phe  
50

<210> 129  
<211> 62  
<212> PRT  
<213> Homo sapien

<400> 129

Met Val Thr Thr Lys Glu Asn Met Tyr Ser Gln Arg Arg Met Arg Lys  
1 5 10 15

Glu Ala Thr Phe Val Thr Thr His Lys Thr Thr Asn His Lys Arg Gln  
20 25 30

His Lys Trp Arg Glu Leu Gln Gly Lys Ala Ile Arg Cys Lys Pro Ser  
35 40 45

Ser Ser Thr Leu Arg Ala Leu Ile Val Met Arg Ala Arg His  
50 55 60

<210> 130  
<211> 38  
<212> PRT  
<213> Homo sapien

<400> 130

Met Ser His His Asn Cys Ala Asn Lys His Ser Cys Val Lys Asn Glu

67

1

5

10

15

Asp Thr Val Phe Tyr Phe Lys Lys Val Gln Tyr Asn Ile Pro Cys Pro  
 20 25 30

Leu Asn Val Glu Ser Phe  
 35

<210> 131  
 <211> 25  
 <212> PRT  
 <213> Homo sapien

<400> 131

Met Arg Arg Ile Leu Ile Asn Gln Lys Lys Cys Tyr Gly Pro Leu Ile  
 1 5 10 15

Glu Met Leu Phe Phe Cys Thr Ser Asn  
 20 25

<210> 132  
 <211> 316  
 <212> PRT  
 <213> Homo sapien

<400> 132

Ile Arg Asn Asp Lys Gly Asp Ile Ala Thr Asp Pro Thr Glu Val Gln  
 1 5 10 15

Thr Ile Ile Arg Glu Tyr Tyr Lys Tyr Leu Tyr Ala Ser Lys Leu Glu  
 20 25 30

Asn Leu Gly Glu Met Asp Lys Phe Met Thr Tyr Thr Leu Pro Arg Leu  
 35 40 45

Lys Gln Glu Glu Ile Glu Ser Leu Lys Arg Pro Ile Ser Cys Ser Glu  
 50 55 60

Ile Glu Ser Val Ile Asn Ser Leu Pro Thr Thr Lys Ser Pro Gly Pro  
 65 70 75 80

Asp Gly Phe Thr Ala Glu Phe Tyr Gln Val Tyr Lys Glu Glu Leu Val  
 85 90 95

Pro Phe Leu Leu Lys Leu Phe Gln Lys Lys Lys Lys Lys Asn Trp Gly  
 100 105 110

Lys Arg Leu Phe Leu Pro Asn Ser Phe Leu Ala Asn Pro Phe Ser Pro  
 115 120 125

Leu Glu Leu Pro Lys Ser Gln Ala Arg Asn Thr Leu Gln Lys Lys Asn  
 130 135 140

Leu Gln Val Ile Met Phe Ser Asn Ala Pro Ile Arg Ile Val Lys Ile  
 145 150 155 160

Leu Leu Leu Arg Lys Asn Tyr Leu Ala Lys Thr Gln Tyr Leu Arg Ile  
 165 170 175

Asn His His Ser Lys Gln Gly Leu Val Leu Leu Ile His Tyr Arg Cys  
 180 185 190

Gly Ile Tyr Tyr Ser Pro Gly Gly Arg Gln Gly Tyr Ala Val Pro Gly  
 195 200 205

Ile Ser Thr Lys Phe Thr Ala Arg Val Val Ile Thr Phe Thr Ile Ile  
 210 215 220

Thr Gly Thr Tyr Lys Asp Lys Asn Pro Met Ala Val Ile Pro Gln Leu  
 225 230 235 240

Asp Val Gln Lys Lys Ser Ile Ser Ile Lys Gly Pro Ala His Phe Phe  
 245 250 255

Ala Leu Ile Lys Ile Leu Leu Ile Gln Ile Leu Ser Gln Ile Ala Gly  
 260 265 270

Phe Asn Gly Lys Thr Pro Ser Gln Lys Leu Arg Ala Ile Tyr Asn Lys  
 275 280 285

Pro Ala Ser Gln Gly Ala Ser Leu Gly Gly Arg His Ala Glu Lys Phe  
 290 295 300

Pro Tyr Thr Ser Gly Val Arg Gln Arg Ala Pro Ile  
 305 310 315

<210> 133  
 <211> 34  
 <212> PRT  
 <213> Homo sapien



69

&lt;400&gt; 133

Met Ala Phe Arg Ile Val Leu Thr Arg Leu Arg Ile Ile Tyr Phe Leu  
1 5 10 15

Leu His His Val Leu Ser Tyr Lys Glu Asp Lys Met Leu Ile Ala Ile  
20 25 30

Gly Asn

&lt;210&gt; 134

&lt;211&gt; 123

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 134

Gln Glu Ala Leu Ala Arg Ile Ala Cys Gln Asn Asn Met Thr Arg His  
1 5 10 15

His Ser Tyr Arg Ser Val Arg Gly Asn Ala Leu Glu Lys Lys Ser Asn  
20 25 30

Tyr Glu Val Leu Glu Lys Asp Val Gly Leu Lys Arg Phe Leu Pro Lys  
35 40 45

Ser Leu Leu Asp Ser Val Arg Ala Lys Thr Leu Arg Lys Leu Met Gln  
50 55 60

Gln Thr Cys Arg Gln Val Thr Asn Leu Asn Arg Glu Glu Ser Ile Leu  
65 70 75 80

Lys Phe Phe Glu Ile Leu Ser Pro Val Tyr Arg Phe Asp Lys Glu Cys  
85 90 95

Phe Lys Cys Ala Leu Gly Ser Ser Trp Ile Ile Ser Val Glu Leu Ala  
100 105 110

Ile Gly Pro Glu Glu Gly Ile Ser Tyr Leu Thr  
115 120

&lt;210&gt; 135

&lt;211&gt; 56

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 135

70

Met Leu Val Thr Ile Phe Tyr Leu Ile Leu Lys Ser Ser Gly Ile Ile  
 1 5 10 15

Met Ser Ile Tyr Leu Ile Leu Gly Met Phe Gln Ile His Phe Gln Glu  
 20 25 30

Trp Val Ser His Ser Leu Phe Thr Tyr Cys Ile Gln Ile Ile Leu Asp  
 35 40 45

Leu Ile Ile Ser Lys Ile His Ile  
 50 55

<210> 136  
 <211> 38  
 <212> PRT  
 <213> Homo sapien

<400> 136

Met Cys Ile Cys Ile Ser Asn Cys Tyr Val Phe Leu Ile Val Asn Leu  
 1 5 10 15

Phe Asn His Cys Lys Met Thr Phe Phe Ile Leu Ser Asn Met Asn Cys  
 20 25 30

Ser Lys Ile Tyr Phe Phe  
 35

<210> 137  
 <211> 30  
 <212> PRT  
 <213> Homo sapien

<400> 137

Met Arg Thr Asn Ile Val Leu Thr Arg Tyr Met Val Leu Arg Ser Val  
 1 5 10 15

Ile Phe Asn Thr Asn Val Leu His Cys Tyr Ser Ile Tyr Leu  
 20 25 30

<210> 138  
 <211> 52  
 <212> PRT  
 <213> Homo sapien

<400> 138

Met Phe Gln Gln Lys Leu Thr Gln Glu Gly Lys Lys Ser Gln Lys His

71

1

5

10

15

Ile Ile Asn Asn Thr Val Cys Asn Leu Ile Ile His Asn Glu Asn Ile  
 20 25 30

Asn His Leu Asn Asn Glu Thr Leu Leu Cys Asn Pro Ile Ile Leu Ile  
 35 40 45

Asn Lys Ile Leu  
 50

<210> 139  
 <211> 70  
 <212> PRT  
 <213> Homo sapien

<400> 139

Met Gly Ser Cys Cys Ser Ser Gln Tyr Val Val Lys Leu Asn Glu Tyr  
 1 5 10 15

Ile Arg His Gly Thr Cys Asn Cys Gly Asn Ala Glu Leu Gln Gly Met  
 20 25 30

His Ile Leu Lys Phe Asn Gly Tyr His Gln Ile Ala Phe His Ile Ile  
 35 40 45

Lys Ile Leu Asn Tyr Lys Gln Glu Asn Thr Ile Met Asp His Ser Asn  
 50 55 60

Gln Glu Asn Phe Phe Phe  
 65 70

<210> 140  
 <211> 52  
 <212> PRT  
 <213> Homo sapien

<400> 140

Met Thr Leu Leu Asn Phe Tyr Phe Arg Phe Arg Gly Ala Cys Val Met  
 1 5 10 15

Ala Val Tyr Cys Lys Pro Tyr Ser Ala Asp Thr Thr Leu Ser Thr Gly  
 20 25 30

Gly Pro Leu Asp His Ala Ser Ile Ser Pro Arg Arg Ile Val Cys Thr  
 35 40 45

Val Ser Ser Glu  
50

<210> 141  
<211> 13  
<212> PRT  
<213> Homo sapien

<400> 141

Met Lys Ala Pro Gly Lys Gln Phe Tyr Ser Asn Arg Ser  
1 5 10

<210> 142  
<211> 54  
<212> PRT  
<213> Homo sapien

<400> 142

Met Phe Trp Ile Pro Val Pro Tyr Thr Val Arg Cys Phe Tyr Lys Tyr  
1 5 10 15

Phe Leu Leu Val Cys Arg Leu Ser Phe His Ser Leu Asn Ser Ile Leu  
20 25 30

Phe Pro Glu Pro Glu Phe Ile Tyr Ser Phe Val Phe Arg Gly Ser Arg  
35 40 45

Ser Val Thr Gln Ala Gly  
50

<210> 143  
<211> 69  
<212> PRT  
<213> Homo sapien

<400> 143

Glu Leu Ala Glu His Phe Val Cys Phe Gly Tyr Gln Ser Leu Ile Gln  
1 5 10 15

Leu Gly Val Phe Ile Asn Ile Phe Ser Ala Ser Val Ala Cys Leu Phe  
20 25 30

Ile Leu Leu Thr Val His Phe Thr Ala Gln Phe Leu Ile Leu Met Lys  
35 40 45

73

Ser Thr Leu Ser Ile Phe Ser Phe Met Asn Tyr Ala Phe Gly Val Leu  
50 55 60

Ser Glu Asn Ser Leu  
65

<210> 144  
<211> 40  
<212> PRT  
<213> Homo sapien

<400> 144

Met Pro Ala Cys Met Tyr Thr Arg Leu Arg Thr Pro Asn Pro Lys Thr  
1 5 10 15

Ile His Cys Ile Glu Cys Val Val Phe Gln Phe Phe Cys Thr Ser Ala  
20 25 30

Ile Leu His Leu Gln His Thr Ala  
35 40

<210> 145  
<211> 35  
<212> PRT  
<213> Homo sapien

<400> 145

Met Lys Gln Ala Lys Lys Lys Lys Lys Arg Lys Glu Arg Lys Lys Lys  
1 5 10 15

Lys Glu Arg Glu Arg Gly Arg Glu Glu Gly Gly Arg Lys Lys Glu Arg  
20 25 30

Gly Gly Arg  
35

<210> 146  
<211> 46  
<212> PRT  
<213> Homo sapien

<400> 146

Met Cys Ile Pro Glu Lys Thr Gly His Phe Ile Gln Asp Gln Glu His  
1 5 10 15

Pro Thr Lys Lys Gln Lys Gln Arg Glu Ile Ser Phe Val Phe Val Ser  
20 25 30

Gln Phe Lys Thr Arg Asn Asn Met Pro Ala Tyr Gly Phe Ser  
           35                          40                          45

<210> 147  
 <211> 45  
 <212> PRT  
 <213> Homo sapien

<400> 147

Met Phe Gln Lys Lys Ser Arg Gly Ser Gln Ile Ser Leu Lys Lys Tyr  
   1                          5                          10                          15

Phe Thr Thr Tyr Phe Phe Ser Gln Ile Cys His Met Glu Leu Cys Ile  
                           20                          25                          30

Ile Ile His Met Asn Ser Gln Phe Ile Thr Tyr Leu Leu  
           35                          40                          45

<210> 148  
 <211> 70  
 <212> PRT  
 <213> Homo sapien

<400> 148

Met Ala Phe Tyr Leu Ile Met Leu Ile Lys Thr Leu Lys Ala Lys His  
   1                          5                          10                          15

Phe Glu Ala Leu Glu Asn Leu Ser Thr Asn Tyr Ala Arg Val Tyr Tyr  
                           20                          25                          30

Lys Leu Ile Ile Lys Asp Thr Ile Val Thr Ala Arg Gly Gly Ala Arg  
           35                          40                          45

Lys Pro Asn Leu Ala Ile Ser Ser His Gly Gly Arg Arg Ala Ala Leu  
   50                          55                          60

Glu Gly Pro Leu Pro Ile  
   65                          70

<210> 149  
 <211> 104  
 <212> PRT  
 <213> Homo sapien

<400> 149

75

Arg Cys Gly Asn Gln Val His Glu Thr Asn Pro Leu Glu Met Leu Arg  
 1 5 10 15

Leu Asp Asn Thr Leu Glu Glu Ile Ile Phe Lys Leu Val Pro Gly Leu  
 20 25 30

Arg Glu Gln Glu Leu Glu Arg Glu Ser Glu Phe Trp Lys Lys Asn Lys  
 35 40 45

Pro Gln Glu Asn Gly Gln Asp Asp Thr Ser Lys Ala Asp Lys Pro Lys  
 50 55 60

Val Asp Glu Glu Gly Asp Glu Asn Glu Asp Asp Lys Asp Tyr His Arg  
 65 70 75 80

Ser Asp Pro Gln Ile Ala Ile Cys Leu Asp Cys Leu Arg Asn Asn Gly  
 85 90 95

Gln Ser Gly Asp Asn Val Val Lys  
 100

&lt;210&gt; 150

&lt;211&gt; 50

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 150

Met Ser Leu Tyr Leu Glu Lys Lys Ser Asn Asn Thr Thr Ser Val Asn  
 1 5 10 15

Phe Cys Ser Ser Glu Lys Ser Ile Ser Ile Thr Pro Val Gly Ser Ser  
 20 25 30

Arg Ser Tyr Ile Pro Pro Leu Ala Lys Val Arg Leu Ile Lys Leu Trp  
 35 40 45

Gly Gly  
 50

&lt;210&gt; 151

&lt;211&gt; 54

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 151

Met Val Leu Leu Ser Ser Ala Met Ser Ser Gln Ile Phe Ser Leu Leu

1

5

10

15

Ile Asp Ser Ser Ile Cys Pro Cys Ser Ser Phe Ser Phe Cys Ser Thr  
35 40 45

Tyr Phe Tyr Ala Ile Leu  
50

<210>	152
<211>	26
<212>	PRT
<213>	Homo sapien

<400> 152

Met Leu Pro Met Ser Leu Arg Arg Tyr His His Tyr Asn Tyr Ser Leu  
1 5 10 15

Ser Trp Tyr Gln Trp Lys Val Asn Leu Thr  
20 25

<210>	153
<211>	36
<212>	PRT
<213>	Homo sapien

<400> 153

Met Gly Gln Ile Lys Ser Leu Gly Ser Asp Asp Gln Met Thr Arg Ser  
1 5 10 15

Ile Cys Lys Thr Ile Leu Asn Phe Gly Glu Ser Phe Pro Ile Phe Thr  
20 25 30

Ala Trp Ile Pro  
35

```
<210> 154
<211> 49
<212> PRT
<213> Homo sapien
```

<400> 154

Met Ser Pro Leu Val Asn Trp Ser Lys Pro Asn Lys Leu Pro Thr Ile  
1 5 10 15



Lys Pro Thr Ser Asn Pro Cys Pro Ser Leu Pro Phe Phe Ala Phe Phe  
 20 25 30

Asn Gly Lys Glu His Lys Arg Arg Ile Gly Cys Leu Phe Ile Ser Phe  
 35 40 45

Phe

<210> 155  
 <211> 54  
 <212> PRT  
 <213> Homo sapien

<400> 155

Met Ser Gln Lys Val Thr Arg Thr Pro Lys Val Val Glu Asn Leu Ile  
 1 5 10 15

Asn Arg His Asn Asn Pro Lys Met Ser Trp Asn Cys Ser Lys Lys Met  
 20 25 30

Gln Thr Ser Gln Leu Gln Gly Asn Phe Arg Asn Asn Arg Ser Asn Phe  
 35 40 45

Gln Arg Ser Ser Ser His  
 50

<210> 156  
 <211> 72  
 <212> PRT  
 <213> Homo sapien

<400> 156

Tyr Ile Leu Asn Phe Phe Tyr Ala Phe Leu Cys Val Val Tyr His Val  
 1 5 10 15

Phe Ser Arg Ile Ser Leu Asn Phe Tyr Tyr Tyr Tyr Tyr Leu Asp Thr  
 20 25 30

Val Ser His Tyr Val Ala Gln Gly Gly Leu Glu Leu Leu Gly Ser Ser  
 35 40 45

Asn Pro Pro Thr Ser Ala Ser His Val Ala Gly Thr Thr Gly Met Tyr  
 50 55 60

78

Leu Cys Leu Val Phe Ser Ala Leu  
65 70

<210> 157  
<211> 69  
<212> PRT  
<213> Homo sapien

<400> 157

Met Asp Leu Arg Thr His Phe Leu Asp Gln Ile Asn Leu Glu Asn Ala  
1 5 10 15

Ile Leu Met Pro Ser Tyr Leu Arg Thr Val Ile Tyr His Phe Asn Ser  
20 25 30

Phe Ser Ala Met Ser His Met Gly Arg Thr Lys His Leu Leu Thr Asn  
35 40 45

Lys Arg Asp Ser Glu Arg Lys Leu Lys Ser Glu Ile Leu Val Glu Lys  
50 55 60

His Ser Lys Arg Ile  
65

<210> 158  
<211> 46  
<212> PRT  
<213> Homo sapien

<400> 158

Met Ser Ser Leu Ala Ala Thr Gln Thr Arg Lys Pro Trp Glu Phe Pro  
1 5 10 15

Ser Ala Val Val Gln Arg Arg Tyr Arg Asn Val Thr Leu His Leu Ile  
20 25 30

Val Thr Cys Ser Val Asn Arg Ile Ala Ser Thr Leu Ala Pro  
35 40 45

<210> 159  
<211> 62  
<212> PRT  
<213> Homo sapien

<400> 159

Met Gln Asn Glu Ser Leu Gln Gly Lys Gln Gly Ile Gln Lys Arg Asn  
1 5 10 15

Lys Asn Cys Lys Met Phe Ser Cys Gln Arg Thr Tyr Lys Lys Leu Ser  
20 25 30

Glu Thr Leu Arg Phe Lys Phe Leu Val Leu Glu Ser Arg Ser Glu Asp  
35 40 45

Pro Gly Glu Arg Glu Lys Gly Val Leu Ser Ile Gln Ile Met  
50 55 60

<210> 160  
<211> 46  
<212> PRT  
<213> Homo sapien

<400> 160

Met Tyr Glu Thr Pro Val His Pro Asp His Asn Pro Thr Phe Leu Thr  
1 5 10 15

Cys Ala Tyr Asn Asn Tyr Leu Ile Ser Asn Met Ser Gln Phe Ser Ile  
20 25 30

Ser Phe Leu Leu Thr Asn Phe Asn Pro Glu Asn Ser Lys Glu  
35 40 45

<210> 161  
<211> 25  
<212> PRT  
<213> Homo sapien

<400> 161

Met Leu Pro Arg Ala Ser Ile Leu Gln Arg Val Leu Phe Lys Asp Tyr  
1 5 10 15

Gly Arg Pro Gln Asp Trp Phe Ile Ile  
20 25

<210> 162  
<211> 33  
<212> PRT  
<213> Homo sapien

<400> 162

Met Leu Ser Thr Gly Ile Leu Ile Leu Ser Leu Gln Lys Ile Asn His  
1 5 10 15

80

Gln Asn His Trp Ile Gln Ile Lys Ile Lys Thr Asn Ser Ala Gln Tyr  
20 25 30

Gly

<210> 163  
<211> 77  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 163

Met Gly Arg Gly Gln Asn Gln Arg Lys Gly Trp Cys Val Ala Thr Val  
1 5 10 15

Leu Gly Met Gly Ala Val Ser Leu Thr Thr Pro Pro Phe Ala Gly Gln  
20 25 30

Glu Cys Ile Cys Phe Ser Gly Ala Arg Pro Arg Pro Cys Arg Phe Arg  
35 40 45

Cys Glu Phe Trp Pro Leu Gly Arg Pro Pro Gly Gly Arg Thr Cys Phe  
50 55 60

Phe Gly His Cys Leu Leu Asn Arg Ala Gln Met Ala Met  
65 70 75

<210> 164  
<211> 34  
<212> PRT  
<213> Homo sapien

&lt;400&gt; 164

Met Ser Thr Ile Ser Ser Ser Pro Leu Pro Asp Ser His Gly Val Thr  
1 5 10 15

His Arg Pro Arg Arg Lys Gly Asn Ser Leu Ile Val Leu Gln Ile Arg  
20 25 30

Asn Gly

<210> 165  
<211> 67  
<212> PRT  
<213> Homo sapien

81

&lt;400&gt; 165

Met Gly Thr Thr Trp Ile Thr Ser Pro Ala Pro Met Gly Trp Asn Ser  
 1 5 10 15

Leu Tyr Arg Val Pro Pro Arg Gly Thr Gln Met Gly Arg Pro Ser Ser  
 20 25 30

Gly Arg Thr Phe Arg Leu Leu Ser Thr Leu Ala Leu Met Asn Asn Ala  
 35 40 45

Ser Met Asn Asn His Ile Gln Val Phe Leu Gly Lys Lys Lys Val Ile  
 50 55 60

Ser Leu Glu  
 65

&lt;210&gt; 166

&lt;211&gt; 46

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 166

Met Gly Leu Tyr Ile Ile Lys Ile Thr Gln Gly Leu Lys Asn Thr Leu  
 1 5 10 15

Asn Pro Trp Phe Leu Leu Ser Val Ile Lys His Ser Leu Ser Lys Leu  
 20 25 30

Ala Cys Val Asn Ala Ile Asn Ile Phe Gln Phe Lys Cys Tyr  
 35 40 45

&lt;210&gt; 167

&lt;211&gt; 54

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 167

Met Cys Thr Ala Arg Gly Lys Trp Phe Tyr Thr Leu Val Ser Trp Val  
 1 5 10 15

Ser Lys Leu Phe Val Gln Thr Leu Ile Cys Phe Leu Glu Lys Val Ala  
 20 25 30

Asp Lys Pro Ile Trp Lys Met Glu Ile Phe Ile Asn Trp Val Asn Leu  
 35 40 45

Val Gly Ile Asp Pro Leu  
50

<210> 168  
<211> 53  
<212> PRT  
<213> Homo sapien

<400> 168

Met His Ser His Phe Tyr Tyr Phe Ile Leu Tyr Gln Tyr Ile Val Phe  
1 5 10 15

Ile Thr Tyr Tyr Tyr Ile Gln Val Phe Leu Leu Ser Ile Leu Ser Arg  
20 25 30

Arg Thr Leu Thr Phe Leu Val Val Glu Gly Leu Arg Ile Arg Ser Glu  
35 40 45

Tyr Leu Glu Ala Lys  
50

<210> 169  
<211> 37  
<212> PRT  
<213> Homo sapien

<400> 169

Met Lys Ser Gly Trp Pro Trp Ser Cys Phe Val Asp Ile Phe Ser Glu  
1 5 10 15

His Ser Ser Ser Ser Trp Ser Pro Cys Arg Lys His Leu Lys Ser Ser  
20 25 30

Lys Leu Asn Lys Ile  
35

<210> 170  
<211> 135  
<212> PRT  
<213> Homo sapien

<400> 170

Met Leu Pro Thr Ile Trp Gly Ala Val Phe Pro Pro Leu Ile Trp Ala  
1 5 10 15

Pro Phe Ile Phe Pro Gly Val Pro His Ile Leu Gln Gly Glu His Pro

83

20

25

30

Ile Gly Pro Lys Pro Cys Ala Ala Thr Ser Pro Phe Pro Tyr Thr Ile  
 35 40 45

Phe Ser Pro Ala Val Lys Phe Asn Pro Phe Ser Pro Pro Pro Arg Phe  
 50 55 60

Ser Gly Tyr Phe Pro Asp Val Pro Pro Pro Phe Leu Arg Ala Ile Pro  
 65 70 75 80

Arg Ser Gly Leu Pro Pro Pro Arg Gly Tyr Ser Pro His Ser Arg Lys  
 85 90 95

Gly Ser Pro His Ile Phe Leu Thr Pro Arg Val Tyr Phe Lys Asn Phe  
 100 105 110

Pro Arg Ile Trp Gly Ala Leu Leu Leu Lys Pro Glu Asn Leu Leu  
 115 120 125

Leu Tyr Gly Gly Pro Leu Ser  
 130 135

<210> 171  
 <211> 57  
 <212> PRT  
 <213> Homo sapien

<400> 171

Met Leu Ile Phe Phe Ser Leu Pro Leu Ala Val Ser Val Thr Met Ser  
 1 5 10 15

Thr Phe Leu Asp Met Phe Ala His Ile Val Leu Pro Ala Glu Thr Glu  
 20 25 30

Asp Leu Gly Leu Gly Leu Ser Ala Leu His Thr Val Pro Ala Cys Ser  
 35 40 45

Pro Val Pro Ser Trp Ile Arg Cys Leu  
 50 55

<210> 172  
 <211> 77  
 <212> PRT  
 <213> Homo sapien

84

&lt;400&gt; 172

Met Glu Gly Tyr Trp Ile Tyr Asn Asn Arg His Ile Ser Lys Val Tyr  
 1 5 10 15

Asn Leu Arg Phe Tyr Ile Met Val Tyr Thr Pro Trp Lys Pro Leu Lys  
 20 25 30

Ile Gly Glu Tyr Ile His His Tyr Ser Pro Lys Ile Phe Leu Met Asn  
 35 40 45

Ser Phe Val Ile Ser Leu Pro Phe Phe Pro Ile Ser Arg Thr Leu Ala  
 50 55 60

Ser Ser Gly Asn His Gly Ser Ala Phe Ser Leu Tyr Arg  
 65 70 75

&lt;210&gt; 173

&lt;211&gt; 33

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 173

Met Met Cys Gln Lys Leu Thr Asp Glu Leu Ile Tyr Ser Val Leu Ser  
 1 5 10 15

Lys Pro Asp Gly Ala Ser Pro Ala Pro Ile Arg Ile Ala Ala His Cys  
 20 25 30

Ala

&lt;210&gt; 174

&lt;211&gt; 48

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 174

Met Thr Glu His Ser Thr Gly Arg Phe Val Trp Tyr Pro Ser Cys Asp  
 1 5 10 15

Glu Ser Asp His Ile Ser Pro Pro Ile Cys Trp Glu Phe Ala Leu Ala  
 20 25 30

Gly Gln Lys Met Trp Thr Gly Ile Ala Thr Thr Ala Leu Gln Pro Gly  
 35 40 45



85

<210> 175  
<211> 57  
<212> PRT  
<213> Homo sapien

<400> 175

Met Ile Leu Asn Ser Leu Ile Ser Pro Leu Gly Leu Ala Leu Ala Lys  
1 5 10 15

Ile Phe Asp Asn Val Ser Gln Asp Ile Leu Arg Asn Asn Thr Lys Lys  
20 25 30

Tyr Gly Leu Asp Ala Asn Ala Ile Lys Val Glu Arg Lys Cys Leu Tyr  
35 40 45

Tyr His Thr Glu Lys Leu Leu Ile Cys  
50 55

<210> 176  
<211> 41  
<212> PRT  
<213> Homo sapien

<400> 176

Met Ile Thr Ile Leu Val His Leu Val Asn Asp Thr Arg Ala Val Leu  
1 5 10 15

Gly Val Pro Gly Lys Gly Ile Pro Glu Ala Gly Lys Leu Thr Ser Thr  
20 25 30

Arg Gly Leu Phe Gly His His Gly Ile  
35 40

<210> 177  
<211> 75  
<212> PRT  
<213> Homo sapien

<400> 177

Met Arg Phe Cys Cys Cys His Phe Ser Thr Val Thr Leu Gly Leu Val  
1 5 10 15

Val Trp Leu Gly Asn Glu Phe Leu Gln Asn Tyr Glu Gly Ile Ala Thr  
20 25 30

Trp Ser Ser Ser Phe Leu Thr Leu Leu Trp Arg Met Arg Ser Leu Lys

86

35

40

45

Pro Phe Asn Ser Leu Ser Phe Leu Gly Asp Phe Ser Pro Ala Leu Asn  
 50 55 60

Cys Leu Val Phe Gln Cys Ser Glu Asn Cys Lys  
 65 70 75

<210> 178  
 <211> 87  
 <212> PRT  
 <213> Homo sapien

<400> 178

Met Val Ile Ile Lys Ile Val Lys Leu Ile Ser Cys Trp Trp Pro Gly  
 1 5 10 15

Ala Val Pro His Ala Cys Ile Pro Ala Leu Cys Asp Ala Glu Ala Gly  
 20 25 30

Met Ile Thr Met Val Arg Met Ile Gly Asp His Pro Val Pro Thr Thr  
 35 40 45

Ser Asp Asn Pro Val Leu Leu Leu Asn Asn Thr Lys Lys Lys Leu Ala  
 50 55 60

Gly Ser Leu Val Val Gly Ile Leu Val Ser Ser His Ala Tyr Pro Arg  
 65 70 75 80

Ser Ala Glu Ala Val Ile Tyr  
 85

<210> 179  
 <211> 541  
 <212> PRT  
 <213> Homo sapien

<400> 179

Met Asp Gly Ala Val Met Glu Gly Pro Leu Phe Leu Gln Ser Gln Arg  
 1 5 10 15

Phe Gly Thr Lys Val Val Trp Arg Met Asp Ala Glu Pro Tyr Pro Gly  
 20 25 30

Ala Ala Trp Val Arg Glu Pro Arg Asn Arg Glu Arg Arg Trp Arg Lys  
 35 40 45

Thr Trp Ala Val Leu Tyr Pro Ala Ser Pro His Gly Val Ala Arg Leu  
 50 55 60

Glu Phe Phe Asp His Lys Gly Ser Ser Ser Gly Gly Gly Arg Gly Ser  
 65 70 75 80

Ser Arg Arg Leu Asp Cys Lys Val Ile Arg Leu Ala Glu Cys Val Ser  
 85 90 95

Val Ala Pro Val Thr Val Glu Thr Pro Pro Glu Pro Gly Ala Thr Ala  
 100 105 110

Phe Arg Leu Asp Thr Ala Gln Arg Ser His Leu Leu Ala Ala Asp Ala  
 115 120 125

Pro Ser Ser Ala Ala Trp Val Gln Thr Leu Cys Arg Asn Ala Phe Pro  
 130 135 140

Lys Gly Ser Trp Thr Leu Ala Pro Thr Asp Asn Pro Pro Lys Leu Ser  
 145 150 155 160

Ala Leu Glu Met Leu Glu Asn Ser Leu Tyr Ser Pro Thr Trp Glu Gly  
 165 170 175

Arg Arg Leu Arg Ser Pro Gly Arg Asp Gly Val Lys Arg Arg Arg Ala  
 180 185 190

Glu Gly Leu Trp Glu Val Gly Gly Tyr Pro Gly Ala His Gly Glu Val  
 195 200 205

Arg Ser Arg Lys Ala Leu Arg Ser Gly Phe Arg Leu Ser Asn Arg Val  
 210 215 220

Cys Leu Pro Gly Ser Gln Phe Trp Val Thr Val Gln Arg Thr Glu Ala  
 225 230 235 240

Ala Glu Arg Cys Gly Leu His Gly Ser Tyr Val Leu Arg Val Glu Ala  
 245 250 255

Glu Arg Leu Thr Leu Leu Thr Val Gly Ala Gln Ser Gln Ile Leu Glu  
 260 265 270

Pro Leu Leu Ser Trp Pro Tyr Thr Leu Leu Arg Arg Tyr Gly Arg Asp

88

275

280

285

Lys Val Met Phe Ser Phe Glu Ala Gly Arg Arg Cys Pro Ser Gly Pro  
 290 295 300

Gly Thr Phe Thr Phe Gln Thr Ala Gln Gly Asn Asp Ile Phe Gln Ala  
 305 310 315 320

Val Glu Thr Ala Ile His Arg Gln Lys Ala Gln Gly Lys Ala Gly Gln  
 325 330 335

Gly His Asp Val Leu Arg Ala Asp Ser His Glu Gly Glu Val Ala Glu  
 340 345 350

Gly Lys Leu Pro Ser Pro Pro Gly Pro Gln Glu Leu Leu Asp Ser Pro  
 355 360 365

Pro Ala Leu Tyr Ala Glu Pro Leu Asp Ser Leu Arg Ile Ala Pro Cys  
 370 375 380

Pro Ser Gln Asp Ser Leu Tyr Ser Asp Pro Leu Asp Ser Thr Ser Ala  
 385 390 395 400

Gln Ala Gly Glu Gly Val Gln Arg Lys Lys Pro Leu Tyr Trp Asp Leu  
 405 410 415

Tyr Glu His Ala Gln Gln Gln Leu Leu Lys Ala Lys Leu Thr Asp Pro  
 420 425 430

Lys Glu Asp Pro Ile Tyr Asp Glu Pro Glu Gly Leu Ala Pro Val Pro  
 435 440 445

Pro Gln Gly Leu Tyr Asp Leu Pro Arg Glu Pro Lys Asp Ala Trp Trp  
 450 455 460

Cys Gln Ala Arg Val Lys Glu Glu Gly Tyr Glu Leu Pro Tyr Asn Pro  
 465 470 475 480

Ala Thr Asp Asp Tyr Ala Val Pro Pro Pro Arg Ser Thr Lys Pro Leu  
 485 490 495

Leu Ala Pro Lys Pro Gln Gly Pro Ala Phe Pro Glu Pro Gly Thr Ala  
 500 505 510

89

Thr Gly Ser Gly Ile Lys Ser His Asn Ser Ala Leu Tyr Ser Gln Arg  
 515 520 525

Ile Gln Ile Pro Gly Arg Gly Lys Gly Glu Gly Gly Gly  
 530 535 540

<210> 180  
 <211> 48  
 <212> PRT  
 <213> Homo sapien

<400> 180

Met Ala Lys Tyr Ile Leu Leu Glu Lys Ser Ala Lys Leu Ile Arg Arg  
 1 5 10 15

Ile Tyr Ser Ala Leu Ser Leu Tyr Ile Ser Val Val Leu Ser Ser Lys  
 20 25 30

Ala Ile Trp Gln Asn Asn Glu Tyr Ile Tyr Ser Ser Lys Glu His Asn  
 35 40 45

<210> 181  
 <211> 46  
 <212> PRT  
 <213> Homo sapien

<400> 181

Met Ala Cys Lys Pro Gly Arg Gly Thr Glu Ser Leu Gln Val Lys Pro  
 1 5 10 15

Thr Glu Leu Gln Pro Pro Ala His Ser Thr Ala Trp Ala Thr Glu Gln  
 20 25 30

Lys Ser Val Ser Lys Lys Lys Lys Lys Lys Leu Leu Val Leu  
 35 40 45

<210> 182  
 <211> 79  
 <212> PRT  
 <213> Homo sapien

<400> 182

Met Gln Lys Glu Gly His Arg Arg Leu Asp Ala Ser Pro Ser Phe Leu  
 1 5 10 15

Gln Glu Leu Leu Ser Glu Asn Asn Thr Lys His Thr Leu Gln His Thr  
 20 25 30

Ser Gly Val Cys Arg Ser Ser Val Ser Ser Cys Phe Gly Ser Leu Cys  
65 70 75 80

91

Tyr Leu Leu Ser Pro Cys Asp Pro Asn Val Arg Ser Leu Phe Leu Tyr  
                     85                    90                    95

Phe Ile Phe Phe Phe Leu His Thr Thr Val Tyr Gly Cys Gln Ile Asp  
                     100                    105                    110

Lys Gly

<210> 185  
 <211> 47  
 <212> PRT  
 <213> Homo sapien

<400> 185

Met Thr Arg Leu Glu Phe His Trp Ser Asn His Gly Ser Leu His Pro  
   1                    5                    10                    15

Arg Pro His Gln Phe Gln Glu Ile Leu Pro Pro Gln Gly Ser Arg Glu  
                     20                    25                    30

Ala Lys Ile Ile Gly Thr Cys Pro Gly Gly Ala Arg Lys Pro Asn  
                     35                    40                    45

<210> 186  
 <211> 82  
 <212> PRT  
 <213> Homo sapien

<400> 186

Met Asn Thr Ser Leu Asp Cys Lys Arg Arg Gln Gly Gln Cys Arg Glu  
   1                    5                    10                    15

His Cys Lys Lys Thr His Arg His Pro Pro Trp Pro Pro Leu Ile Ser  
                     20                    25                    30

Ala Val Ala Thr Ser Gly Gln Val Ala Pro Ile Gly Ala Gln Met Leu  
                     35                    40                    45

Leu Ser Leu Thr Ala Ile Leu Ile Val His Glu Val Ala Cys Ser Ser  
                     50                    55                    60

Ala Phe Pro Pro Gln Ala Arg Ser Pro Ala Pro Met Glu His His Lys  
                     65                    70                    75                    80

Ser Val

<210> 187  
 <211> 85  
 <212> PRT  
 <213> Homo sapien  
 <400> 187

Met Glu Phe Gly Phe Glu Arg Pro Pro Gly Gln Val Pro Leu Lys Leu  
 1 5 10 15

Leu Leu Pro Phe Phe Phe Gly Pro His Leu Asp Arg Leu Thr Arg Lys  
 20 25 30

Pro Met Tyr Ala Ser Ser Ser Ser Ile Cys Glu Lys Phe Lys Leu Cys  
 35 40 45

Lys Ser Ser Thr Cys Thr Trp Glu Leu Phe Phe Ile Pro Thr Leu Tyr  
 50 55 60

Gln Leu Glu Thr Pro Ile Pro Leu His Leu Arg Glu Glu Thr Thr Pro  
 65 70 75 80

Ser Tyr Cys Leu Met  
 85

<210> 188  
 <211> 72  
 <212> PRT  
 <213> Homo sapien

<400> 188

Met Pro Cys His Ser Ile Leu Pro Tyr Tyr Thr Ile Phe Ser Phe Lys  
 1 5 10 15

Gly Phe Ile Phe Pro Thr Ser Leu Ser Leu Lys Gly Arg Ser Gln Asn  
 20 25 30

Ser Cys Met Gly Ile Thr Pro Val Thr Met His Ile Gly Phe Val Ile  
 35 40 45

Asn Ile Ser Glu Lys Ser Asn Met Met Asn Glu Asn Leu Ser Asn Asn  
 50 55 60

Val Asn Lys Ala Tyr Arg Ile Gln



65

70

<210> 189  
 <211> 31  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 189

Met Arg Pro Arg Tyr Asn Asn Leu Phe Ala Leu Phe Phe Leu Pro Leu  
 1 5 10 15

Asn Phe Ser Val Val Ser Leu Ala Met Phe Leu Glu Lys Arg Ser  
 20 25 30

<210> 190  
 <211> 125  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 190

Met Ala Ala Ala Phe Ser Pro Pro Ser Leu Pro Val Pro Ser Leu Leu  
 1 5 10 15

Ser Ser Phe Ser Pro Ser Ala Arg Arg Pro Pro Ala Leu Thr Ser Ser  
 20 25 30

Pro Pro Pro Pro Pro Val Ala Ser Pro Ala Arg Ala Ala Arg Arg Arg  
 35 40 45

Pro Pro Ala Pro Pro Ser Ser His Pro Pro Arg Ala Pro Pro Pro Pro  
 50 55 60

Ser Ser Ser Pro Leu Pro Pro Leu Pro Pro Arg Ala Leu Pro Leu Ser  
 65 70 75 80

Ala Leu Pro Pro Leu Ala Ser Ser Pro Leu Phe Leu Phe Pro Pro Leu  
 85 90 95

Asn Ile Ile Leu Cys Val Trp Arg Asp Ile Leu Phe Val Ser Arg Arg  
 100 105 110

Arg Phe Lys His Thr His Cys Ser His Thr His Gly Arg  
 115 120 125

<210> 191  
 <211> 57

94

<212> PRT  
<213> Homo sapien

<400> 191

Met Ile Leu Lys Leu Leu Gln Gln Leu Tyr Lys Val Thr Gln Asn His  
1 5 10 15

Val Thr Leu Phe Ser Tyr Leu Ser Leu Leu Leu Pro Asp His Cys Gln  
20 25 30

His Asn Phe Tyr Thr Ser Ser Pro Gln Ser Ala Ser Leu Gly His Ala  
35 40 45

Pro Gln Tyr Ala Val Ile Phe Phe Val  
50 55

<210> 192  
<211> 19  
<212> PRT  
<213> Homo sapien

<400> 192

Met Ser Thr Leu Leu Met Asn Pro Ile Lys Cys Thr Pro Tyr Cys Lys  
1 5 10 15

Leu Gln Val

<210> 193  
<211> 33  
<212> PRT  
<213> Homo sapien

<400> 193

Met Arg Lys Ile Tyr Gly Gly His Val Thr Arg Leu Thr Asn Asn Leu  
1 5 10 15

Tyr Cys Pro Gly Gly Ala Arg Lys Pro Asn Ser Ser Thr Leu Arg Ala  
20 25 30

Leu

<210> 194  
<211> 53  
<212> PRT  
<213> Homo sapien

95

&lt;400&gt; 194

Met Ala Trp Leu Ile Phe Phe Val Phe Phe Val Glu Thr Gly Phe His  
 1 5 10 15

His Val Ala Gln Gly Gly Leu Lys Leu Leu Ser Ser Ser Asn Gln Pro  
 20 25 30

Pro Lys Val Phe Gly Ile Thr Gly Ala Thr Tyr Leu Ala Gln Pro Lys  
 35 40 45

Ile Val Phe Val Ser  
 50

&lt;210&gt; 195

&lt;211&gt; 41

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 195

Met Arg Leu Cys Val Ser Met Leu Ile Ser Tyr Leu Ile Lys Arg Arg  
 1 5 10 15

Lys Lys Tyr Ser Pro Glu His Val Ser Arg Phe Gln Ile Ile Ile His  
 20 25 30

Ala Arg Asp Arg Phe Lys Gln Asp Leu  
 35 40

&lt;210&gt; 196

&lt;211&gt; 78

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 196

Met Asn Ser Gln Val Phe Val Leu Ala Cys Pro Arg Pro Ser Tyr Tyr  
 1 5 10 15

Pro Lys Arg Trp Leu Cys Ser Leu Cys Ile Trp Val Thr Ser Thr Lys  
 20 25 30

Ser Ile Ser Asn Tyr Leu Lys His Ser Val Ser Ser Ile Cys Lys Met  
 35 40 45

Arg Ile Asn Asn Val Thr Ser Gln Leu Thr Gly Cys Ser Glu Asp Ser  
 50 55 60

Thr Arg Tyr Cys Ile Gln Ile Thr Ser Val Leu Leu Thr Ser  
65 70 75

<210> 197  
<211> 38  
<212> PRT  
<213> Homo sapien

<400> 197

Met Leu Ala Leu Ala Gly Val His Leu Pro Gly Ala Ala Arg Lys Pro  
1 5 10 15

Ile Pro Ala His Cys Ala Cys Ile Ser Asp Gly Ala Arg Leu Thr Gly  
20 25 30

Thr Phe Ser Phe Phe Leu  
35

<210> 198  
<211> 27  
<212> PRT  
<213> Homo sapien

<400> 198

Met Gln Thr Glu Lys Val Cys Gln Ser Phe Gly Tyr Val Tyr Val Ile  
1 5 10 15

Ala Tyr Leu Leu Trp Ile Pro Leu Ile Ser Lys  
20 25

<210> 199  
<211> 15  
<212> PRT  
<213> Homo sapien

<400> 199

Met Leu Leu Glu Gly Phe Val Phe Val Leu Leu Leu Lys Leu Trp  
1 5 10 15

<210> 200  
<211> 106  
<212> PRT  
<213> Homo sapien

<400> 200

Met Gly Leu Thr Arg Thr Ser Ala Arg Gln Ser Val Gly Glu Tyr Thr

97

1

5

10

15

Cys Asp Leu Arg Val Val Ile Gly Val Glu Thr Val Arg Gln Pro Gly  
20 25 30

Leu Gln Ile Ala Pro Glu Arg Thr Val Tyr Gln Thr Ala Lys Thr Lys  
35 40 45

Glu Gly Glu Arg Gly Gly Ser Glu Arg Gln Thr Arg Glu Arg Arg Arg  
50 55 60

Arg Glu Arg Glu Glu Arg Arg Arg Asp Glu Glu Ser Gly Glu Gly Thr  
65 70 75 80

Arg Lys Arg Arg Glu Gly Arg Ala Ala Lys Arg Thr Ala Gly Glu Gly  
85 90 95

Gly Arg Arg Gly Gly Glu Ala Thr Arg Glu  
100 105

<210> 201  
<211> 69  
<212> PRT  
<213> Homo sapien

<400> 201

Met Leu Arg Phe Gly Ser Ser Leu Ile Phe Leu Thr Leu Ile Val His  
1 5 10 15

Ile Leu Tyr Leu Ser Leu Gly Ser Cys Asn Arg Met Val Tyr Val Leu  
20 25 30

Lys Ala Thr Leu Arg Lys Phe Ile Ser Tyr Leu Tyr Thr Thr Gly Asp  
35 40 45

Leu Tyr Asn Ser Val Thr Lys Phe Pro Trp Ile Val Gln Lys Asn Gln  
50 55 60

Phe Thr Phe Ser Tyr  
65

<210> 202  
<211> 90  
<212> PRT  
<213> Homo sapien

98

&lt;400&gt; 202

Met Ala Asn Trp Ile Met Leu Met Ile Leu Asn Leu Lys Ile Ser Asn  
 1 5 10 15

Lys Asn Phe Asn Ile His Lys Ala Lys Thr Asp Lys Ala Lys Arg Arg  
 20 25 30

Asn Lys Glu Ile His Asn His Asn Gly Arg Phe Tyr Thr Ser Leu Ser  
 35 40 45

Glu Thr Asp Ile Cys Arg Gln Lys Leu Val Arg Ile Gln Asn Met Leu  
 50 55 60

Thr Gln Leu Asn Lys Met Asp Thr Pro Arg Ala Val Tyr Leu Val Asn  
 65 70 75 80

Ala Leu Leu His Val Leu Tyr Lys Tyr Glu  
 85 90

&lt;210&gt; 203

&lt;211&gt; 65

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 203

Met His Lys Asn Arg Gln Phe Thr Gln Lys Glu Ile His Thr Ser Trp  
 1 5 10 15

Ser Leu Asn Thr Leu Arg Arg Cys Ser Thr Ser Leu Leu Ile Lys Lys  
 20 25 30

Cys Lys Ile Asn Tyr Thr Lys Val Ser Phe Ser Pro Thr Asn Phe Ser  
 35 40 45

Lys Lys Ile Pro Gln Leu Asp Asn Gly Gly Val Ser Tyr Leu Leu Ser  
 50 55 60

Leu  
 65

&lt;210&gt; 204

&lt;211&gt; 34

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 204

99

Met Leu Thr Glu Ser Arg Glu Glu Lys Asn Leu Arg Lys Arg Arg Lys  
1 5 10 15

Leu Asp Phe Trp Phe Phe Glu Thr Ala Gly Lys Lys Gly Gly Phe Gly  
20 25 30

Gly Lys

<210> 205  
<211> 48  
<212> PRT  
<213> Homo sapien

<400> 205

Met Glu His Phe Tyr Ser Cys Gly Asp Ile Gly Phe Tyr Leu Val Asn  
1 5 10 15

Leu Leu Phe Lys Leu Phe Ile Thr Tyr Ser Asp Asn Phe Leu Lys Arg  
20 25 30

Gln Ile Ile Phe Asn Tyr Leu Ile Leu Arg Lys Met Pro Pro His Phe  
35 40 45

<210> 206  
<211> 33  
<212> PRT  
<213> Homo sapien

<400> 206

Met Leu Ile Phe Asn Cys Pro Asn Tyr His Leu Phe Val Phe Leu Thr  
1 5 10 15

Ser Arg Thr Lys Leu Gln Ile Val Ser Ile Thr Asn Phe Tyr Phe Cys  
20 25 30

Lys

<210> 207  
<211> 63  
<212> PRT  
<213> Homo sapien

<400> 207

Met Thr Lys Gln Met Ala Ala Val Glu Thr Ser Phe Pro Pro Leu Pro

**PCT/US01/45177**

```
<210> 208
<211> 32
<212> PRT
<213> Homo sapien
```

Met Ile Val Pro Ala Arg Ala Pro Leu Glu Ser Thr Asn Ser Ser Thr  
1 5 10 15

```
<210> 209
<211> 53
<212> PRT
<213> Homo sapien
```

Met Ser Glu Arg Gly Phe His Gln Gln Lys His Ser Ile Gly Cys Ile  
1 5 10 15

Leu Leu Lys Ile Arg Trp Leu Ile His Asp Leu Leu His Leu Cys Gly  
35 40 45

```
<210> 210
<211> 56
<212> PRT
<213> Homo sapien
```

<400> 210



101

Met Gly Val Ser His Lys Ser Met Gly Lys Ala Leu Ser Pro Thr Phe  
 1 5 10 15

Tyr Phe Phe Leu Phe Ile Tyr Cys Leu Leu Thr Met Tyr Pro Pro  
 20 25 30

Thr Pro Pro Asn Ile Ser Val Thr Phe Lys Gly Ala Ser Thr Phe Leu  
 35 40 45

Phe Thr Ala Val Thr Leu Asn Ala  
 50 55

<210> 211  
 <211> 67  
 <212> PRT  
 <213> Homo sapien

<400> 211

Met Thr Leu Ala Leu Phe Pro Ser Asp Ile Arg Ile Phe Pro Val Lys  
 1 5 10 15

Val Leu Leu Leu Val Asn Ser His Cys Gly Arg Leu Pro Cys Leu Ser  
 20 25 30

Ser Lys Gln Gln Val Cys His Asn Gln Ala Phe Pro Tyr Pro Arg Asn  
 35 40 45

Leu Ser Arg His Ile Ile Ala Gln Phe Gln Ser Pro Thr Ile Ser Pro  
 50 55 60

Phe Leu Pro  
 65

<210> 212  
 <211> 117  
 <212> PRT  
 <213> Homo sapien

<400> 212

Met Leu Cys Asp Arg Arg Glu Thr Ile Ser His Gln Ala Thr Ala Phe  
 1 5 10 15

Gly Pro Lys Gly Tyr Pro His Asn Cys Gly Asp Gln Asn Ser Gly Asp  
 20 25 30

102

Pro Leu Ser Val Pro Gly Arg Pro Met Gly Arg Trp Lys Ser Arg Leu  
                   35                  40                  45

Lys Arg Leu Val Ala Arg Pro Glu Gly Ala Pro Asn Thr Gly Arg Gln  
           50                  55                  60

Arg Pro Leu Arg Ala Asn Pro Gly Ala Gln His Ala Phe Asp Val Gln  
   65                  70                  75                  80

Lys Asp Phe Phe Ser Ala Gln Ile Leu Leu Val Gly Gly Gly Tyr Asn  
                   85                  90                  95

Trp Lys Ile Asp Gly Thr Lys His Leu Phe Cys Phe Tyr Lys Ala Ser  
                   100                  105                  110

Ile Gln Leu Ile His  
           115

<210> 213  
 <211> 39  
 <212> PRT  
 <213> Homo sapien

<400> 213

Met Ala Ala Asn Asn Phe Ser Gly Leu Gly Asp Glu Arg Leu Ser Cys  
   1                  5                  10                  15

Gln Thr Gly Gln Ile Glu Arg His Thr Thr Phe Trp Gln Leu Ile Tyr  
           20                  25                  30

Phe Leu Phe Ile Leu Phe Tyr  
           35

<210> 214  
 <211> 48  
 <212> PRT  
 <213> Homo sapien

<400> 214

Met Asp Ala Phe Leu Val Ile Ile Cys Tyr Lys Lys Pro Ser Pro Lys  
   1                  5                  10                  15

Ile Asn Asn Met Pro Glu Cys Ser His Phe Tyr Leu Leu Tyr Ala Arg  
           20                  25                  30

Glu Ala Pro Val Ile Thr Lys Thr His Cys Pro Cys Pro Arg Ile Lys

103

35

40

45

<210> 215  
 <211> 23  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 215

Met Ile Gly Lys Ile Thr Arg Val Val Glu Lys Lys Thr Leu Gly Leu  
 1 5 10 15

Val Ser Val Pro Gln Lys Ser  
 20

<210> 216  
 <211> 49  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 216

Met Leu Arg Val Lys Asn Trp Glu Ile Gln Thr Gln Ile Leu Leu Arg  
 1 5 10 15

Leu Asp Gln Ser Ile Phe Ile Lys Cys Leu Val Gly His Lys Asn Thr  
 20 25 30

Pro Ile Thr Glu Leu Ala Tyr Tyr Tyr Pro Leu Tyr Asn Ser Arg Glu  
 35 40 45

Ser

<210> 217  
 <211> 89  
 <212> PRT  
 <213> Homo sapien

&lt;400&gt; 217

Met Arg Leu Ile Ile Cys Thr Ser Val Asp Trp Asn Asn Ser Ile Ile  
 1 5 10 15

Ser Leu Pro Asn Val Glu Trp Met Pro His Pro Ile Leu Leu Lys Phe  
 20 25 30

Cys Asn Ser Asn Arg Ile Ala Asn Ile Asn Ile Phe Phe Leu Ser Cys  
 35 40 45

104

Asn Ala Trp Thr Val Phe Glu Ala Leu Gly His Trp Phe Phe Ser Val  
50 55 60

Pro Phe Phe Phe Ile Phe Leu Phe Leu Gly Gly Glu Glu Ser Phe Phe  
65 70 75 80

Ser Lys Thr Lys Gln Lys Gly Leu Leu  
85

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number  
**WO 02/042329 A3**

(51) International Patent Classification<sup>7</sup>: **C07K 14/47**,  
16/18, C12Q 1/68, G01N 33/577, A61K 38/17, 31/7088

(US). SUN, Yongming [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(21) International Application Number: PCT/US01/45177

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date:  
20 November 2001 (20.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/252,189 21 November 2000 (21.11.2000) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/252,189 (CIP)  
Filed on 21 November 2000 (21.11.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): DI-ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

Published:  
— with international search report

(72) Inventors; and  
(75) Inventors/Applicants (*for US only*): SALCEDA, Susana [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). MACINA, Roberto, A. [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). RECIPON, Herve [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115

(88) Date of publication of the international search report:  
26 June 2003

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/042329 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate tissue, identifying prostate tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

## INTERNATIONAL SEARCH REPORT

 Int. Application No  
 PCT/US 01/45177

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C07K14/47 C07K16/18 C12Q1/68 G01N33/577 A61K38/17 A61K31/7088		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, EMBL, SEQUENCE SEARCH		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 60162 A (ALI SHUJATH ; DIADEXUS LLC (US); SALCEDA SUSANA (US); SUN YONGMING) 25 November 1999 (1999-11-25)	
A	WO 00 14234 A (UROCOR INC) 16 March 2000 (2000-03-16)	
A	WO 98 35031 A (UNIV ROCHESTER MEDICAL CENTER ; BARTH RICHARD K (US); WEI CHUNGWEN) 13 August 1998 (1998-08-13)	
A	WO 00 58470 A (MOUL JUDD W ; SRIVASTAVA SHIV (US); ZOU ZHIQIANG (US); SRIKANTAN VA) 5 October 2000 (2000-10-05)	
<input type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search  18 November 2002		Date of mailing of the international search report  12.03.03
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  Espen, J

# INTERNATIONAL SEARCH REPORT

ational application No.  
PCT/US 01/45177

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-17 in part

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1; Claims: in part: 1-17; all as far as applicable

Nucleic acid sequence relating to SEQ ID No 1 and amino acid sequence relating to SEQ ID NO 115. Kit comprising said sequences and vaccine relating to said sequences. Methods for determining the presence of a prostate specific nucleic acid/protein in a sample. Method for diagnosing and monitoring the presence of metastases of prostate cancer in a patient by making use of said sequences.

Inventions 2-114; Claims: in part: 1-17; all as far as applicable

as invention 1 but limited to subject-matter relating to SEQ ID NOs 2-114 and SEQ ID NOs 116-217

invention 2 is limited to SEQ ID NOs 2 and 116

invention 3 is limited to SEQ ID Nos 3 and 117, etc. ...

invention 114 is limited to SEQ ID NOs 114 and 217



# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No  
PCT/US 01/45177

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9960162	A	25-11-1999	CA 2328377 A1 EP 1082460 A1 JP 2002515263 T WO 9960162 A1	25-11-1999 14-03-2001 28-05-2002 25-11-1999
WO 0014234	A	16-03-2000	AU 5815699 A WO 0014234 A1	27-03-2000 16-03-2000
WO 9835031	A	13-08-1998	US 6100444 A AU 6269098 A EP 0983352 A1 JP 2001511014 T WO 9835031 A1	08-08-2000 26-08-1998 08-03-2000 07-08-2001 13-08-1998
WO 0058470	A	05-10-2000	AU 3919600 A EP 1185639 A1 WO 0058470 A1	16-10-2000 13-03-2002 05-10-2000